SAICSIT2024

SOUTH AFRICAN INSTITUTE OF COMPUTER SCIENTISTS
AND INFORMATION TECHNOLOGISTS

SAICSIT'24

Online Proceedings of the South African Institute of Computer Scientists and Information Technologists 2024 Conference

ISSN 2959-8877

Human-Machine-Digital-Convergence

The South African Institute of Computer Scientists and Information Technologists (SAICSIT[1]) 2024 conference was hosted by the Computer Sciences Department and the School of Information Technology at Nelson Mandela University. The conference took place in Gqeberha from Monday the 15th to Wednesday the 17th of July 2024.

SAICSIT has been holding this annual conference since 1985. The Institute focuses on research and development in computing and information technology (IT) in Southern Africa. SAICSIT 2024 is the 45th annual conference and will accommodate two tracks, Computer Science and Information Systems, with various topics considered for each track. The conference theme, Human-Machine-Digital-Convergence, aims to stimulate robust engagement that will generate new insights into the transforming relationship between 'humans and machines'.

[1] https://saicsit.org/

SAICSIT2024

SOUTH AFRICAN INSTITUTE OF COMPUTER SCIENTISTS
AND INFORMATION TECHNOLOGISTS

Preface

This volume contains the accepted and revised research papers accepted for the online proceedings of SAICSIT 2024, the *45th Conference of the South African Institute of Computer Scientists and Information Technologists*. SAICSIT 2024 was held from 15 to 17 July 2024 in the Boardwalk Hotel, Gqeberha, South Africa. The theme of the conference, *Human-Machine-Digital-Convergence* is aimed at stimulating robust engagement that will generate new insights into the transforming relationship between humans and machines.

The role of the 45th SAICSIT conference is to support and connect the Southern African computing community so that innovation and creativity can flourish. The SAICSIT conference aims to provide a space to meet and exchange ideas on addressing the challenges of the fast-evolving digital future, and we are therefore glad to report that submissions were received from authors with more than 60 unique national and international affiliations.

The review process was double-blind and rigorous, with every paper sent to at least three and receiving at least two substantive reviews from our program committee. The program committee comprised more than 125 members with over 50 national and international affiliations. The technical committee, comprising of the technical chairs and track chairs, managed the review process and supported the authors in revising their papers to the quality results that are published in this proceedings.

This SAICSIT 2024 online proceedings book contains 20 full research papers organised in their respective Computer Science and Information Systems research tracks. These submissions were included for presentation in the SAICSIT 2024 programme. With more than 90 papers submitted (from national and international authors) and more than 80 submissions in total sent out for review, the acceptance rate for full research papers for this online volume is 25%. In the Information Systems track, 38 submissions were sent out for review, and 10 were accepted for this volume, amounting to an acceptance rate of 26%. In the Computer Science track, 48 submissions were sent out for review, and 10 were accepted for this volume, amounting to an acceptance rate of 20.8%.

We want to express our gratitude to the track chairs and program committee reviewers for their hard work and dedication. Thank you to the authors of all the submitted papers for sharing their research results. We hope the opportunity to participate in SAICSIT 2024 will have a lasting impact on the quality and productivity of future research in our scholarly community.

We also acknowledge the enthusiasm and outstanding contributions of the local organisers of SAICSIT 2024, the Computer Sciences Department and the School of Information Technology at Nelson Mandela University. Thank you to everyone who contributed to the success of SAICSIT 2024.

June 2024

Aurona Gerber
Technical Chair: SAICSIT 2024

SAICSIT 2024 Sponsors

The sponsors of SAICSIT 2024 are herewith gratefully acknowledged.

wirk
we create software
DSA
BATSAMAYI
aws
VSC
SMART SUPPLY CHAINS
avochoc
TANGIBLE
START
DemystifyCoding

SAICSIT 2024 Organisation

General Chair

Mathys C. du Plessis	Nelson Mandela University, South Africa

Technical Chair

Aurona Gerber	University of the Western Cape and CAIR, South Africa

Computer Science Co-Chairs

Hein Venter	University of Pretoria, South Africa
Lester Cowley	Nelson Mandela University, South Africa

Information Systems Co-Chairs

Marie Hattingh	University of Pretoria, South Africa
Melissa Makalima	Nelson Mandela University, South Africa

Publication Chair

Aurona Gerber	University of the Western Cape and CAIR, South Africa

SAICSIT 2024 Programme Committee

Computer Science Track

Adedayo, Oluwasola Mary	University of Winnipeg, Canada
Adigun, Matthew	University of Zululand, South Africa

Bagula, Antoine	University of the Western Cape, South Africa
Bradshaw, Karen	Rhodes University, South Africa
Casini, Giovanni	ISTI - CNR, Italy
Chavula, Josiah	University of Cape Town, South Africa
Chindipha, Stones	Rhodes University, South Africa
Dlamini, Moses T.	CSIR, Pretoria, South Africa
Du Plessis, M.C.	Nelson Mandela University, South Africa
Dunaiski, Marcel	Stellenbosch University,], South Africa
Furnell, Steven	University of Nottingham, United Kingdom
Greyling, Jean	Nelson Mandela University, South Africa
Grobler, Trienko	Stellenbosch University, South Africa
Gruner, Stefan	University of Pretoria, South Africa
Hazelhurst, Scott	University of the Witwatersrand, South Africa
Henney, Andre	University of the Western Cape, South Africa
Hutchison, Andrew	Google Switzerland, Switzerland
Isafiade, Omowunmi Elizabeth	University of the Western Cape, South Africa
James, Steven	University of the Witwatersrand, South Africa
Jembere, Edgar	University of KwaZulu-Natal, South Africa
Klein, Richard	University of the Witwatersrand, South Africa
Kotzé, Eduan	University of the Free State, South Africa
Kuttel, Michelle	University of Cape Town, South Africa
Leenen, Louise	University of the Western Cape, South Africa
Makura, Sheunesu	University of Pretoria, South Africa
Marais, Patrick	University of Cape Town, South Africa
Medupe, Abiodun	University of Pretoria, South Africa
Meyer, Thomas	University of Cape Town and CAIR, South Africa
Modipa, Thipe	University of Limpopo, South Africa
Mokwena, Sello	University of Limpopo, South Africa
Motara, Yusuf	Rhodes University, South Africa
Nel, Stephan	Stellenbosch University, South Africa
Ngoqo, Bukelwa	Nelson Mandela University, South Africa
Norman, Michael	University of the Western Cape, South Africa
Nyathi, Thambo	University of Pretoria, South Africa
Nyirenda, Clement	University of the Western Cape, South Africa
Olivier, Martin	University of Pretoria, South Africa
Rananga, Seani	University of Pretoria, South Africa
Sanders, Ian	University of South Africa (UNISA), South Africa
Serfontein, Rudi	North-West University, South Africa
Shibeshi, Zelalem	Rhodes University, South Africa
Singh, Avinash	University of Pretoria, South Africa
Suleman, Hussein	University of Cape Town, South Africa
Timm, Nils	University of Pretoria, South Africa
Vadapalli, Hima Bindu	University of Johannesburg, South Africa
van Alten, Clint	University of the Witwatersrand, South Africa
van der Merwe, Brink	Stellenbosch University, South Africa

Van Zijl, Lynette	Stellenbosch University, South Africa
Van Heerden, Willem	University of Pretoria, South Africa
Velempini, Mthulisi	University of Limpopo, South Africa
Venter, H.S.	University of Pretoria, Pretoria, South Africa
Venter, Isabella	University of the Western Cape, South Africa
Wa Nkongolo, Nkongolo	University of Pretoria, South Africa
Watson, Bruce	Stellenbosch University, South Africa
Zugenmaier, Alf	Hochschule München, Germany

Information Systems Track

Adebesin, Funmi	University of Pretoria, South Africa
Adeliyi, Timothy	University of Pretoria, South Africa
Baduza, Gugulethu	Rhodes University, South Africa
Beelders, Tanya	University of the Free State, South Africa
Bottomley, Edward-John	Stellenbosch University, South Africa
Campher, Susanna E. S.	North-West University, South Africa
Chindenga, Edmore	University of Fort Hare, South Africa
Davids, Zane	University of Cape Town, South Africa
De Wet, Lizette	University of the Free State, South Africa
Du Plessis, MC	Nelson Mandela University, South Africa
du Preez, Madely	University of Pretoria, South Africa
Eybers, Sunet	University of South Africa (UNISA), South Africa
Harmse, Rudi	Nelson Mandela University, South Africa
Hattingh, Marié	University of Pretoria, South Africa
James, Steven	University of the Witwatersrand, South Africa
Jere, Nobert	Walter Sisulu University, South Africa
Le Roux, Daniel	Stellenbosch University, South Africa
Maasdorp, Christiaan	Stellenbosch University, South Africa
Makalima, Melissa	Nelson Mandela University, South Africa
Masinde, Muthoni	Central University of Technology, South Africa
Matthee, Machdel	University of Pretoria, South Africa
Mawela, Tendani	University of Pretoria, South Africa
Mennega, Nita	University of Pretoria, South Africa
Mohammed, Nouralden	University of the Witwatersrand, South Africa
Mujinga, Mathias	University of South Africa (UNISA), South Africa
Mwansa, Gardner	Walter Sisulu University, South Africa
Ncube, Zenzo Polite	University of Mpumalanga, South Africa
Nel, Wynand	University of the Free State, South Africae
Ngoqo, Bukelwa	Nelson Mandela University, South Africa
Oki, Olukayode	Walter Sisulu University, South Africa
Olaitan, Oo	Walter Sisulu University, South Africa
Oluwadele, Deborah	University of Pretoria, South Africa
Parry, Douglas	Stellenbosch University, South Africa

Pillay, Komla	University of Johannesburg, South Africa
Pretorius, Henk	University of Pretoria, South Africa
Serfontein, Rudi	North-West University, South Africa
Seymour, Lisa	University of Cape Town, South Africa
Sigama, Khuliso	Tshwane University of Technology, South Africa
Smit, Imelda	North-West University, South Africa
Smit, Danie	University of Pretoria, South Africa
Smuts, Hanlie	University of Pretoria, South Africa
Snyman, Dirk	University of Cape Town, South Africa
Steyn, Riana	University of Pretoria, South Africa
Taylor, Estelle	North-West University, South Africa
Tuyikeze, Tite	Sol Plaatje University, South Africa
van der Merwe, Thomas	University of South Africa (UNISA), South Africa
van der Vyver, Charles	North-West University, South Africa
van Eck, Rene	Vaal University of Technology, South Africa
Van Staden, Corné	University of South Africa (UNISA), South Africa
Wa Nkongolo, Nkongolo	University of Pretoria, South Africa
Weilbach, Lizette	University of Pretoria, South Africa

Postgraduate Symposium

Wesson, Janet	Nelson Mandela University, South Africa

Table of Contents

II Information Systems Track 149

Part I

Computer Science Track

Defining an Ontology for Ant-like Robots Based on Simulated Ant-agent Characteristics

Colin Chibaya

Dept of Computer Science & Information Technology
Sol Plaatje University

colin.chibaya@spu.ac.za

Abstract. Swarm intelligence systems, where robotic devices are programmed with distinct abilities that operate at the individual level to produce collective emergent behaviour, are particularly promising in fields like nanotechnology. These systems are typically employed to solve complex real-world problems at minimal costs. For instance, ant colony systems emulate the behaviour of natural ants to tackle challenging issues. Solutions to complex optimization problems, such as the bridge crossing problem, vehicle routing problems, shortest path formation problem, and the travelling salesman problem, have been developed using this approach. This study draws inspiration from various simulated ant colony systems, exploring the low-level actions and capabilities of simulated ant-like robotic devices to develop an ant-bots swarm intelligence ontology. An ant-bot is conceived as a small, simple autonomous robot modelled after simulated ants. Individually, an ant-bot may not accomplish much, but as part of a swarm, these robots can generate impressive emergent behaviour. We examine the specific aspects of simulated ant agents that lead to emergent behaviour and incorporate these into the design of an ant-bots swarm intelligence ontology. Experimental tests identified three key components as the foundation of the desired swarm intelligence ontology. First, the swarm space component captures metadata about the configuration of the simulated environments, targets, and any global swarm rules. Second, the ant-bot. context emphasizes the individual abilities and activities of ant-bots. Lastly, the swarm interaction component details the communication mechanisms used, whether direct or indirect, local or global, inspired by nature, mathematical models, biological processes, or other methods. This swarm intelligence ontology serves as a formal knowledge representation model for ant-bots, encapsulating these aspects to enable effective swarm emergent behaviour.

Keywords: Ant system; ant-bot; ant-bot ontology

1 Introduction

Swarm intelligence involves designing intelligent multi-agent systems that emulate the collective natural behaviours of social colonies like ants, termites, birds, spiders, and bees [1]. This relatively recent problem-solving approach draws inspiration from nature [2]. However, the exact methods governing the behaviour of natural colonies are not fully understood. It is evident that cooperative colony behaviour arises from simple interactions between colony members, but what specific actions do individual members take to generate collective behavior?

Ants, in particular, are intriguing, and related ant systems have garnered significant attention due to their practical applications in combinatorial optimization problems such as shortest path formation [4][5]. But what exactly do individual simulated ant agents do to achieve swarm convergence? How can we formalize and represent the knowledge embedded in ant systems to enhance their practical application in real-world problem-solving?

Ontology, a branch of philosophy, explores concepts of existence, being, becoming, and reality [6]. Therefore, a swarm intelligence ontology refers to a structured collection of knowledge about swarm behaviour, abilities of swarm members, the environment in which the swarm operates, and other relevant parameters. This study aims to design a swarm intelligence ontology inspired by ant systems to better understand and utilize their principles in solving practical problems.

1.1 Problem definition

Creating an ant-bots ontology involves coordinating homogeneous swarms. Extending this effort to coordinate heterogeneous swarms is an ambitious project that requires understanding various distinct swarm intelligence ontologies. For instance, we would need separate ontologies for ants, termites, bees, social spiders, fish, or birds before developing a generic swarm intelligence ontology capable of handling heterogeneity.

Simulated ant systems are particularly fascinating. What are the fundamental elements of an ant-bots ontology that could contribute to the advancement of heterogeneity in swarm intelligence? There has been little discussion about the abilities of ant-bots to create emergent behaviour. The individual actions of ant-bots, their communication and interaction strategies, decision-making processes, and the information they generate are all critical components of swarm knowledge, yet their representation remains undefined. This study aims to identify and characterize the key aspects of simulated ant systems that define an ant-bots ontology.

To achieve this, we need to conceptualize the environment in which ant-bots operate, design the ant-bots themselves, and understand the processes through which ant-bots interact, communicate, share, or create knowledge. Ultimately, we seek to demonstrate a comprehensive understanding of the prospective vocabulary of ant-bots to enhance the application and visibility of ant systems. To the best of our knowledge, formalizing the representation of ant systems in the form of an ontology is a novel and significant contribution to the field.

1.2 Overview

The article is organized as follows: Section 2 reviews the literature, identifying key aspects of ant systems that could inspire the design of an ant-bots ontology. Section 3 characterizes these identified aspects, bringing us closer to defining the component units of the envisioned ant-bots ontology. In Section 4, we propose the actual ant-bots ontology. Finally, Section 5 concludes the study, highlighting our contributions and suggesting directions for future research.

2. Relate Work

Research efforts to represent knowledge in a substantive and methodological manner are evident [8]. Although there have been attempts in the literature to create swarm intelligence ontologies inspired by ant colony systems, bee colony optimization models, particle swarm optimization systems, and hybrid models combining these approaches, the specific and explicit representation of ant-bots knowledge remains unclear. This study aims to investigate specific aspects of simulated ant systems to propose a clear ant-bots ontology.

The closest work in the literature involves studies on the role of pheromones in guiding ant-like agents moving between two points [2]. These studies established that pheromones act as indirect guides towards agents' targets, with ants probabilistically following paths marked by pheromones. The direction selection by ant-bots, referred to as orientation, is based on the pheromone levels around a decision-making agent. This understanding of pheromones as guides has been refined and optimized over time [9]. Despite the consistent concept of pheromone perception for orientation, heuristic information is necessary to achieve optimal solutions [10]. For example, defining swarm memory as being held in the environment is a heuristic feature that helps reduce the cost of managing ant-bots.

Most ant systems imply that ant agents possess an internal state to keep track of the swarm goal [19]. Each agent is aware of its target, understanding which pheromone levels are attractive or repulsive [19]. Switching between different internal states is a reward mechanism when an agent finds the target or successfully returns to the nest [19]. During the search, ant agents drop, and update pheromone levels based on their internal state [19]. Pheromone updates are also managed heuristically through dissipation processes to optimize swarm convergence quality and speed [19].

In summary, most ant systems achieve mission planning and execution through:

(a) pheromone management policies, including detecting pheromone levels, dropping new pheromone levels, updating pheromone quantities, and allow for pheromone dissipation.

(b) internal state transitions, including managing context awareness and switching internal states as necessary.

(c) local search procedures such as orientation and movement.

A theoretical survey on ant colony optimization has emphasized understanding the theoretical basis of swarm convergence on optimal solutions driven by stochastic methods [11]. Trust level computation, as demonstrated in related studies [12], highlights the value of these swarm aspects for recommending an ant-bots ontology [13]. The primary goal of this study is to explicitly elucidate the key aspects of ant systems that may inspire the design of an ant-bots ontology, facilitating the practical application of related swarm intelligence systems in real life.

3. Methods and materials

This study is in its early stages toward creating heterogeneous ontologies. It focuses on understanding a specific homogeneous ant-bot ontology, which will later be integrated with other homogeneous ontologies to address heterogeneity in swarm intelligence models. Specifically, we undertake (a) a requirements elicitation exercise (identifying key aspects in the design of an ant-bots ontology), (b) requirements specification (determining how each identified aspect fits into the problem), and (c) ontology modeling and proposing a methodology for integrating these aspects.

Design science research is utilized to define the main computational artifacts of the study [14]. In proposing the ant-bots ontology, emphasis is placed on achieving scalability, reproducibility, and adaptability. A positivist approach drives the study, aiming to verify the work through deductive methods until an ant-bots ontology emerges [15][16][17]. Ideally, the results should be transferable from practical insights to theoretical frameworks [18]. The next three subsections provide detailed analyses of the key aspects commonly considered in studies of ant systems.

3.1 The ant-bot architecture

An ant-bot is designed with basic memory to hold four key pieces of information: (a) its position, (b) its internal state, (c) neighborhood, and available instruction set, as illustrated in Figure 1. Positional awareness allows an ant-bot to retrieve and update the levels of pheromone at its location. This is an individual property of the ant-bot. The internal state maintains the ant-bot’s role within the swarm. At any time, an ant-bot is either searching for the target or returning to the nest. In each state, it deposits a different type of pheromone at its current position, updating the swarm’s global information. Typically, an ant-bot places pheromones that attract other ant-bots in the opposite state. For instance, an ant-bot searching for the target releases pheromones that attract ant-bots returning to the nest, and vice versa, leading to the emergence of well-trodden paths at the swarm level.

Each ant-bot is also equipped to perceive its neighborhood, which includes potential locations to move to. These locations are weighted based on their attractiveness or repulsiveness, determined by the levels of specific pheromones they contain. An ant-bot decides stochastically where to move next, considering the attractiveness and repulsiveness of the surrounding locations.

Finally, ant-bots follow deterministic instructions to achieve their goals. These instructions, executed at the individual level, lead to emergent behavior at the swarm level. The collection of these instructions and their parameters is discussed further in the next section on ant-bots' actions. This work does not delve into the engineering details of ant-bot design.

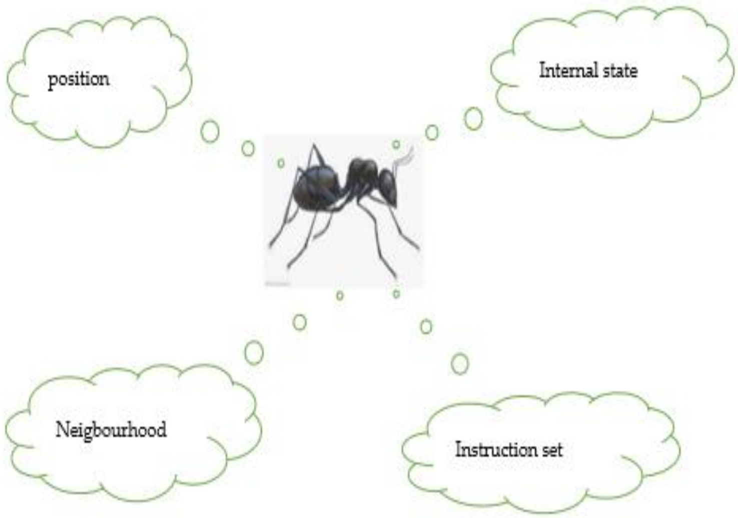


Fig. 1. Ant-bot design

3.2 The ant-bot actions

It is evident that ant-bots deposit distinct levels of pheromone as they traverse the environment. This act of depositing pheromone modifies the amounts of this particular level of pheromone present at the ant-bot’s current position. The quantity of pheromone dispersed in the environment serves as a collective memory for the swarm. Utilizing the environment to store these quantities offers the advantage of reducing the memory burden on individual ant-bots. Additionally, it separates the existence of ant-bots from the overall swarm solution. When an ant-bot deposits and updates specific pheromone levels at its current location, this action simultaneously updates the shared memory at the swarm level. Listing 1 provides a computational representation of the pheromone deposition concept.

Listing 1: dropping specific levels of pheromone

```
Drop (int x, int y, int Qty)
{
            update (level, Qty + read(level, x, y))
}
```

In mathematical notation, Listing 1 can be expressed as: P(x, y) = P(x,y) + Q, where P (x,y) is the pheromone level at position (x,y). Q is the quantity of pheromone to be added. Thus, the function Drop (x, y, Qty) updates the pheromone level at the specified coordinates by adding the specified quantity to the current level.

The subsequent task for an ant-bot involves relocating after depositing specific levels of pheromone at its current position. This relocation process necessitates the ant-bot to initially orient itself before proceeding in the chosen direction. Orientation, from a computational standpoint, involves the ant-bot assessing the attractiveness or repulsiveness of all neighboring locations before selecting a direction to move towards. This assessment is accomplished by assigning weights to each neighboring location based on the levels of attractive and repulsive pheromones it contains. Consequently, locations with higher levels of attractive pheromones are favored, while those with higher levels of repulsive pheromones are penalized. Listing 2 outlines the orientation concept algorithmically.

Listing 2: ant-bot orientation

```
Orientate (int x, int y)
{
     for each nearbyLocation i
    {
    Wi←qtyAtt(x±[0,1];y±[0,1])–qtyRep(x±[0,1];y±[0,1])
     }
    set a scaled roulette wheel for wi
    direction ← randPick (i,wi)
}
```

The mathematical notation suggests that Orientate (x,y) represents the function to orientate the ant-bot at coordinates (x,y). The function qtyAtt(x±[0,1],y±[0,1]) and qtyRep(x±[0,1],y±[0,1]) denote the quantities of attractive and repulsive pheromones, respectively, at nearby locations around (x,y). Wi is the weight assigned to each nearby location, calculated as the difference between the quantities of attractive and repulsive pheromones. Also, randPick (i,Wi) selects a random nearby location based on the scaled roulette wheel determined by Wi. The formula within the loop in Listing 2 outlines the process for calculating the weight of each nearby location surrounding an ant-bot. Subsequently, a stochastic roulette wheel is constructed based on these weights, where locations with higher attractiveness are assigned wider spans compared to repulsive locations. In this setup, random selection of a location tends to favor attractive locations over repulsive ones. An advantage of this approach is that even highly repulsive locations have a chance of being randomly selected, introducing an element of randomness in movement. However, it's uncommon for a repulsive location to be chosen. After successful orientation, ant-bot movement ensues. Here, movement involves relocating from the current location to the destination determined by the orientation direction. The concept of movement is elaborated in Listing 3.

Listing 3: ant-bot movement

```
Move (int x, int y, direction i)
{
        x ← x + direction ($i_x$)
        y ← y + direction ($i_y$)
}
```

In this scenario, the ant-bot will shift from its current x-coordinate towards the x-direction indicated by the orientation. Similarly, it will transition from its y-coordinate towards the y-direction specified by the orientation. Eventually, as a moving ant-bot progresses, it will reach its target. This event prompts the transition to the flip state action. If an ant-bot was previously searching for food, it will now alter its role within the swarm to begin seeking the nest. Consequently, its perception of attractive and repulsive pheromones will change, along with a reversal in the levels of pheromones deposited at each visited location, reflecting opposite behaviors in all instances. Listing 4 elucidates the process of transitioning between different internal states.

Listing 4: ant-bot flip state

```
state Flip (int x, int y)
{
        if (x ; y) has Target
                return homing
        if (x ; y) has Nest
                return searching
}
```

Specifically, this aspect conditionally transitions an ant-bot to a particular internal state based on its current location. If an ant-bot finds itself at the target because it was previously searching for it, it switches to "homing," initiating the journey back to the nest. Conversely, if an ant-bot arrives at the nest while homing, it switches back to "searching" mode and recommences search trips. This function exclusively affects the two target locations; otherwise, a searching ant-bot continues its search uninterrupted. An ant-bot traveling towards the nest but not yet arrived will persist in homing until it reaches the nest. This function encompasses all ant-bot actions outlined in the ant-bots instruction set depicted in Figure 1 previously.

3.3 The ant-bot's heuristic aspects

In ant-bot systems, two heuristic elements are implied. Evidently, swarms of ant-bots are expected to function within deterministic environments, which consist of locations, tuples to store various pheromone levels, and the context in which ant-bots exist or navigate when solving presented problems. We mentioned that environments serve as the shared memory for the swarm. Figure 2 illustrates the environment we envision, delineating locations in a grid format. Each location serves as a habitat for ant-bots and also retains the various pheromone levels previously deposited. The definition of environments is heuristic, established during simulated parameter configuration.

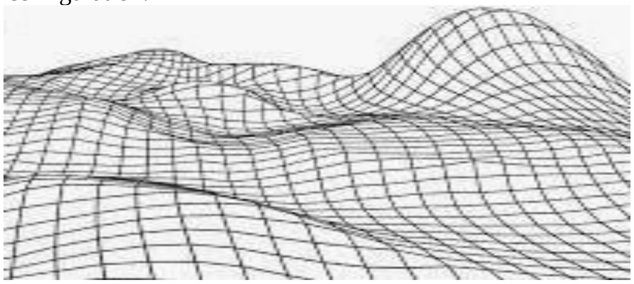


Fig. 2. Ant-bots environments

An intriguing aspect is that pheromone levels deposited at various locations in the environment may gradually dissipate due to evaporation or diffusion. Pheromone dissipation serves as an elitist strategy to refine the paths formed by pheromones and facilitates the process of forgetting old solutions in favor of new ones. While these factors are not directly linked to the individual actions of an ant-bot, they significantly influence ant-bot behaviors, thus warranting inclusion in the design of an ant-bots ontology.

4. Implementation

4.1 Towards the ant-bots Ontology

Figure 3 illustrates the integration of the ant-bots architecture, ant-bots actions, and related heuristic aspects into an ant-bot ontology. This proposed ontology primarily highlights the locations of various pieces of ant-bot knowledge. At its center, the ontology encompasses a comprehensive representation of swarm-level knowledge, which synchronizes information from supporting sub-domains. The heuristic aspects pertain to environmental design and the maintenance and smoothing of pheromone levels in the environment, extending beyond individual ant-bot capabilities.

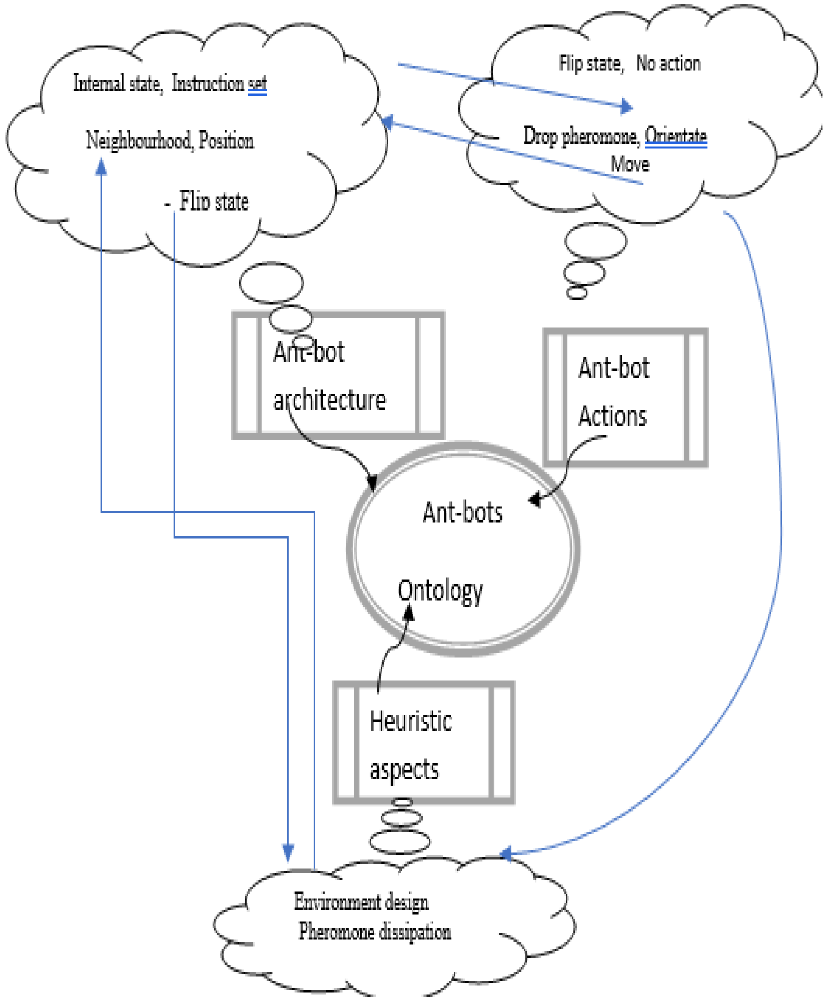


Fig. 3. Ant-bots ontology

The core is also supported by the ant-bot architecture, which encompasses positional knowledge, internal state details (recalling each ant-bot's goal), its neighborhood, and the usage of instructions contained in the knowledge domain. These instructions include actions such as dropping pheromones, moving, orienting, and flipping states.

It is important to note that the survival of an ant-bot swarm relies on the strength of the shared memory, defined by the levels of pheromone present in the environment. Low pheromone levels indicate a lack of swarm-level knowledge about the target's location, leading to random movements. Conversely, excessive pheromone levels can saturate the environment and degrade the established paths. A balance is maintained through elitist heuristic methods, including pheromone dissipation.

This study does not focus on implementation details. Instead, it emphasizes the logic for emulating ant behavior, specifically understanding what each ant-bot does at an individual level to achieve emergent swarm behavior. The focus is on homogeneous swarms of ant-bots, each being simple and autonomous. The ontology implies only stigmergic communication, indirectly mediated through the environment, with no direct messaging between ant-bots. Additionally, an ant-bot can choose to remain inactive when necessary, hence the "No action" aspect in the ant-bot instruction set.

5. Conclusions

This article explored the aspects of ant systems that lead to emergent behavior, interpreting each from a computational perspective. It then proposed a method to integrate these aspects into a knowledge representation framework referred to as an ant-bots ontology. The ant-bots ontology aims to encompass all key aspects necessary for coordinating swarms of homogeneous, ant-like robotic devices. Detailing the elements of the ant-bots ontology is crucial for creating a comprehensive modeling scenario applicable to various swarm formations. The clarity achieved in designing the ant-bots ontology can facilitate application-specific modeling solutions for practical problems. The ant-bots ontology presented in this article includes three interrelated knowledge domains that connect to the core domain. These three aspects and their relational inferences constitute the ontology.

5.1 Contributions

This study makes three key contributions:

- The article proposed a formal knowledge representation approach for defining ant-inspired swarm intelligence systems, thereby extending existing literature in the field.
- It contributes to ongoing efforts to formalize knowledge representation in swarm intelligence models. This study, focused on ant systems, aims to understand the key knowledge domains of swarm intelligence systems, providing insights that can lead to more general applications.

- While the primary focus of this study was on developing an ontology for a homogeneous swarm of ant-bots, this work establishes a foundation for addressing heterogeneity in future research.

5.2 Future work

Four ambitious directions for future work are envisioned from this study:

- Practical experimentation to validate the ant-bot knowledge representation is forthcoming.
- The ant-bot ontology could be enhanced by incorporating additional relevant aspects to broaden its applicability. It should be tested for completeness, optimality, applicability, and expandability.
- Integrating the ant-bot ontology with other swarm intelligence ontologies could eventually lead to the development of practical heterogeneous swarms.
- Extending the ant-bot ontology to handle fuzzy situations, uncertainty, incompleteness, inaccuracies, vagueness, or imprecision is a promising area for future research.

References

1. V Selvi and R Umarani. “Comparative analysis of ant colony and particle swarm optimization techniques”. In: International Journal of Computer Applications 5.4 (2010).
2. Marco Dorigo, Mauro Birattari, and Thomas Stutzle. “Ant colony optimization”. In: IEEE computational intelligence magazine 1.4 (2006).
3. Saurabh Mittal and Larry Rainey. “Harnessing emergence: The control and design of emergent behavior in system of systems engineering”. In: Proceed- ings of the conference on summer computer simulation. 2015.
4. Marco Dorigo and Christian Blum. “Ant colony optimization theory: A sur- vey”. In: Theoretical computer science 344.2-3 (2005).
5. A Kaveh and S Talatahari. “An improved ant colony optimization for con- strained engineering design problems”. In: Engineering Computations (2010).
6. Katie Moon and Deborah Blackman. “A guide to understanding social science research for natural scientists”. In: Conservation biology 28.5 (2014).
7. Jing Wang and Gerardo Beni. “Cellular robotic system with stationary robots and its application to manufacturing lattices”. In: Proceedings. IEEE International Symposium on Intelligent Control 1989. IEEE. 1989.
8. Andrew Booth. “Searching for qualitative research for inclusion in system- atic reviews: a structured methodological review”. In: Systematic reviews 5.1 (2016).
9. JinFeng Wang, XiaoLiang Fan, and Haimin Ding. “An improved ant colony optimization approach for optimization of process planning”. In: The Scientific World Journal 2014.
10. V Maniezzo, LM Gambardella, and FD Luigi. New Optimization Techniques in Engineering, ch. Ant Colony Optimization. 2004.
11. Marco Dorigo and Christian Blum. “Ant colony optimization theory: A sur- vey”. In: Theoretical computer science 344.2-3 (2005).
12. Katie Moon and Deborah Blackman. “A guide to understanding social science research

for natural scientists". In: Conservation biology 28.5 (2014).

13. O Deepa and A Senthilkumar. "Swarm intelligence from natural to artificial systems: Ant colony optimization". In: Networks (Graph-Hoc) 8.1 (2016).
14. V Raghunatha Reddy and A Rajasekhar Reddy. "Lifetime Improvement of WSN by Trust Level based Ant Colony Optimization". In: International Journal of Computer Science and Information Technologies 5.5 (2014).
15. Gale M Sinatra and Doug Lombardi. "Evaluating sources of scientific evidence and claims in the post-truth era may require reappraising plausibility judgments". In: Educational Psychologist 55.3 (2020).
16. Alexander M Novikov and Dmitry A Novikov. Research methodology: From philosophy of science to research design. Vol. 2. CRC Press, 2013.
17. Garima Malhotra. "Strategies in research". In: International Journal for Advance Research and Development 2.5 (2017).
18. Adam DI Kramer, Jamie E Guillory, and Jeffrey T Hancock. "Experimental evidence of massive-scale emotional contagion through social networks". In: Proceedings of the National Academy of Sciences 111.24 (2014).
19. M Bala Varsha, Manoj Kumar, and Neeraj Kumar. "Hybrid TABU-GA search for energy efficient routing In WSN". In: International Journal of Recent Technology and Engineering (IJRTE) 8.4 (2019).

A Technique for the Detection of PDF Tampering or Forgery

Gabriel Grobler, Sheunesu Makura[2][0000-0002-5129-3216] and Hein Venter[3][0000-0002-3607-8630]

University of Pretoria, Hatfield 0028, South Africa
[1]u20534541@tuks.co.za, [2]makura.sm@up.ac.za, [3]hein.venter@up.ac.za

Abstract. Tampering or forgery of digital documents has become widespread, most commonly through altering images without any malicious intent such as enhancing the overall appearance of the image. However, there are occasions when tampering of digital documents can have negative consequences, such as financial fraud and reputational damage. Tampering can occur through altering a digital document's text or editing an image's pixels. Many techniques have been developed to detect whether changes have been made to a document. Most of these techniques rely on generating hashes or watermarking the document. These techniques however have limitations in that they cannot detect alterations to portable document format (PDF) signatures or other non-visual aspects, such as metadata. This paper presents a new technique that can be used to detect tampering within a PDF document by utilizing the PDF document's file page objects. The technique employs a prototype that can detect changes to a PDF document, such as changes made to the text, images, or metadata of the said file.

Keywords: PDF, Tampering, Forgery, Metadata, Hashing, Alterations, File page objects

1 Introduction

Digital media is a common form of communication and entertainment in modern society. Given the widespread use of these digital forms of media through images, videos, and text-based documents, it has opened a new avenue for tampering with official or unofficial content [1]. Forgery is when a person creates a copy of the original, often intending to commit fraud. Tampering is when a person interferes with or changes the original without permission. In the context of this paper, both terminologies will be used interchangeably.

The tampering that the average person will generally encounter is focused on the alteration of images for the purpose of looking better in posted photos or videos through techniques such as deep fakes [2], where a deep fake is a fake image of an event produced using deep learning techniques. Though many alterations are not of a criminal nature there are alterations of more formal digital documents that could have adverse consequences.

One of the most common forms of digital documents used for communications is the portable document format (PDF). A PDF is a file format developed by Adobe that preserves the layout and formatting of a document regardless of the software, hardware, or operating system used to view or print it. PDF files can contain text, images, forms, annotations, and other data, and they are widely used for distributing documents that need to be displayed and printed consistently across different platforms. Given that the PDF is difficult to alter, it is used in many formal forms of communication, such as memorandums, contracts, specifications, etc. However, with the advancements in technology and the expanded ease of access to tools such as Adobe Acrobat or one of the many free online editors that make altering documents in this format possible, it has become simpler to alter PDF files with limited knowledge of the PDF format. This leads to the necessity to detect and analyze any alterations made to a document in PDF format.

Most of the current techniques that are used for the detection and partial analysis of any alterations use a form of watermarking. Though we will highlight two forms of watermarking in this paper, there are numerous ways of implementing such a watermark into a PDF document or most other forms of digital content.

In addition to watermarking, another technique used to check whether the current PDF matches the original is hashing. Though these techniques successfully detect changes to a PDF document and its contents, they mostly rely on the visible aspects of the said PDF, such as the images and text that a reader would see. This is because they mainly rely on using the visible content to the reader. The techniques discussed would not detect changes to this PDF document's metadata or background data. Should alterations be made that would embed malware into the PDF using the scripting abilities of the PDF format, the above techniques would not be able to detect such changes to the document [7]. Nor would they be able to detect changes to a PDF signature, which could have severe consequences should a PDF document be attacked in such a manner [8]. PDF signatures are elaborated in section 2.3.

It is worth noting that while the existing methods of using watermarking and hashing have their limitations, changes can be identified even though the exact point of change may not be identifiable. This is because process of changing anything in a document creates a different hash for the document which therefore makes it difficult to pinpoint exactly where and what caused the change.

1.1 Problem Statement and Research Questions

The problem that this paper is addressing is the tampering or forgery of PDF documents. With the PDF format being used as a formal means of communication in multiple industries, it has become a good target for criminals who wish to affect contracts or aid in misinformation. The adverse effects of such documentation being tampered with or forged could greatly impact many people's lives if it goes undetected. This paper presents a new technique that can detect tampering or forgery of a PDF document using the underlying content of a PDF document, such as the file page objects. File page objects specifically pertain to the content and structure of individual pages within the PDF document.

The main research problem described above can be expanded by asking the following research questions (RQs). These questions are used to assess whether the proposed solution successfully detects tampering or forgery of a PDF document.

RQ.1 What are the current techniques that are used or have been researched for the detection or tampering of PDF documents?
This question aims to address what are the current techniques that are employed for the detection of tampering or forgery in PDF documents. In addition, we would want to find out the current state regarding the detection of tampering or forgery within a PDF document.

RQ.2 Will using the file page objects to generate a hash detect tampering or forgery of a PDF document's text?
This question aims to address whether using the file page objects of a PDF document to generate a hash value will be able to detect tampering or forgery of the document. It applies in two parts: the first using the rewrite approach or the second using the incremental approach. The technique should be successful regardless of how the changes are made to the PDF document.

RQ.3 Can a prototype be developed to detect alterations made to an image within the PDF document?
This question addresses whether the proposed prototype, which is the main contribution of this paper detects alterations made to an image in a PDF document. If the image is altered by methods such as Photoshop or completely replaced, the prototype should be able to detect that a change has been made to such images in the PDF document.

The remainder of the paper is structured as follows: section 2 highlights literature survey of the current techniques used in detecting forgery or tampering, section 3 elaborates on the prosed prototype design and implementation, section 4 discusses the experiments conducted, the results and evaluation then section 5 concludes the paper.

2 Literature Review

This section provides a literature survey of the state of the current research conducted on the topic of detecting tampering or forgery within PDF documents.

2.1 Watermarking

A watermark is a technique used to protect an original piece of work from being copied. When applied to PDF documents, they are not as visible as a watermark on an image; they are generally embedded and hidden from view. Research by Khadam et al. [5] used a watermark on PDF documents using file page objects from a PDF document. The technique was designed to be non-intrusive, whereby the watermark was embedded into the PDF document to be used for later comparison. The watermark was designed to be fragile, meaning it becomes invalid if an attempt is made to alter the document. While this watermarking technique presented by the authors successfully detects changes to a PDF document's format without bloating the file's size, it does not discuss the idea of detecting whether JavaScript has been embedded into the document.

Research by Usop [4] presented a watermarking technique where the PDF file was first converted to an image in the Bitmap (BMP) format, initiating the watermark creation process. After dividing the image pixels into blocks, these blocks were used in a zigzag manner to create watermarks per block. This method of watermarking efficiently detects changes to the visual appearance of the document. However, it cannot detect alterations that do not affect the document's visual appearance to the reader, such as embedding JavaScript code or tampering with the PDF file signature. Another issue is that generating these watermarks can take a long time, but the high detection accuracy justifies the duration.

Dikanev et al. [9] generated a semi-fragile watermark to protect and detect any alterations made to a PDF document. The authors used Quick Response (QR) codes to generate the watermarks. Initially, the page is converted into a raster image, a simple pixel map, which is compressed before being used to generate the QR code. This QR code is then normalized into a specified range before being embedded into the image of the PDF page using an inverse function. When it is time to verify the QR code of the PDF document, the inverse of the process described above is performed. This technique makes it easy to detect changes to the visual appearance of a PDF document and makes it simple to localize where these changes were made. However, should an individual decide to alter the metadata or embed JavaScript code into a PDF document, this technique would be unable to detect such actions.

Research by Jiang et al. [13] presented a technique for embedding encrypted watermarks into various objects within the PDF file, such as text, images, and forms, ensuring resilience against multiple types of attacks like text editing, format modification, and page extraction. The authors used a tamper detection algorithm to verify the integrity of the content and identify any tampered areas. Their technique was tested on various PDF files, demonstrating high performance in terms of imperceptibility, capacity, robustness, and compatibility compared to existing methods [15].

2.2 Hashing

Hashing is an integrity validation technique that relies on one-way hash functions. These functions take an input and produce a string value of what appears to be random characters, which are calculated based on the input.

A hash algorithm was used by Senkyire and Kester [6] as the primary component of their method to detect tampering or forgery of a document. The authors used the SHA-384 algorithm to calculate hashes for the specific input. The SHA-384 algorithm utilizes blocks of size 1024 bits for generating its hashes, which are padded if the data is not the correct size. The process for generating the hash is as follows: the data is divided into n blocks, and the first block is hashed and fed into the next stage of the process. The second block is hashed using the previous block's hash and contents, which are then fed into the following block for hashing. This process repeats until the n-th block. The hash produced by the last block is used for comparison purposes [6]. This method can be applied to a PDF document by converting its pages to images and following the above process.

Our proposed prototype, presented in Section 3, will use hashing to assess PDFs for forgery. The generated hashes can be checked to determine if any changes were made to the PDF.

2.3 PDF Signatures

PDF signatures (or digital signatures) are methods used to validate the origin of a PDF document [8]. They verify the authenticity and integrity of the PDF. For a digital signature to be generated for a PDF document, incremental saving must be used so that the PDF signature can be appended as an object to the document, much like an addition or change to the original document. A single document can have multiple signature objects, allowing it to be signed by multiple stakeholders and verifying that those signatures are valid.

The main concept of the paper by Mladenov et al. [8] revolved around exploiting a PDF document without invalidating the PDF signature. Consequently, changes can be made to the document without the person being able to repudiate it. How this is achieved varies but can broadly fall under the following categories: Universal Signature Forgery, Incremental Saving Attack, or Signature Wrapping Attack [8]. These categories broadly aim to make the PDF signature considered valid. Methods used to protect against such attacks focus solely on the PDF signature and are therefore difficult to expand to apply to the rest of the PDF document. Using a signature may validate the origin of the document, but the technique of using file page objects to generate a hash, allowing detection of tampering and semi-localizing where it happened, will aid in validating the document's integrity.

The concept of shadow attacks on PDF signatures was presented by Mainka et al. [10]. The authors looked at PDF attacks that comply with the PDF standard. While most attacks on PDF signatures rely on creating malformed incremental updates, shadow attacks use well-formed incremental updates. Three forms of shadow attacks are discussed: Hide, Replace, and Hide-and-Replace [10]. A Hide attack attempts to conceal relevant content behind a visible layer, such as text behind an image. The second attack, Replace, uses an incremental update to overwrite the previously declared object. Finally, a Hide-and-Replace attack relies on sending the document to contain hidden descriptions of another document. Once they receive the signed document, they append an Xref table that references this other document.

2.4 Hiding Malicious Content

The premise behind hiding malicious content relies on using a dual file [7]. A dual file is a file that combines two different file formats into a single entity. Popescu [7] used dual file combines a PDF and Tag Image File Format (TIFF) file. TIFF files are generally used for editing and manipulating high-resolution images. When the victim receives the document, they see the PDF version and sign it, relying on the fact that PDF signatures will not be invalidated when this dual file is converted into the TIFF format. Once an attacker converts the file into this format, they can alter the document's contents, and the document will still have a valid PDF signature. This attack is possible because a TIFF header can be placed within the header of a PDF file without causing issues.

Popescu [7] describes what is known as a Dali attack. This attack leverages the concept of a polymorphic file that combines both PDF and TIFF formats. The attacker hides malicious content within the TIFF part of the file. When the file is initially viewed, only the benign PDF content is visible. However, once the digital signature is applied, the file can be manipulated to reveal the hidden malicious content without invalidating the digital signature.

The concepts of embedding content in places not checked by watermarking techniques and other hashing techniques show that using hashing on the file page objects alone will be insufficient to detect all alterations. This therefore prompts the need for new techniques that can be utilized to other unique objects of a PDF document, such as the signature, the header, and possibly the cross-reference table itself.

2.5 Other techniques

Research work by Guangyong et al. [11] presented a novel approach to protecting and tracing the copyright of PDF files using blockchain technology. The authors proposed a blockchain-based copyright protection technique that combines blockchain with data hiding in PDF files. This scheme involves: (i) a new information hiding algorithm that embeds copyright information in PDF files without altering their appearance, (ii) smart contracts for access control and on-chain proofs of PDF file ownership [11]. The experimental evaluation showed that the proposed technique was effective in terms of security and traceability. The data hiding method did not affect the visual quality of the PDF, and the blockchain-based system ensures reliable copyright management.

Research work by Nguinabe et al. [12] introduced a novel approach to detect falsified PDF documents using graph isomorphism. Their technique involved transforming a PDF document into a graph where nodes represent words and edges represent the semantic relationships. The goal was to find an isomorphism between the original and the altered PDF document graphs. Their technique demonstrated 90% accuracy in detecting falsifications, proving its robustness against insertion, deletion, and modification attacks.

3 Methodology and Prototype design

This section discusses the prototype's design and the details of its implementation. An initial discussion regarding the PDF structure is provided before delving into the functionality of the prototype components and their interactions.

3.1 PDF Structure

A PDF document comprises four main sections [3]: the header, body, cross-reference table, and trailer.

- **Header:** Specifies the version number of the PDF format that the document conforms to.
- **Body:** Composed of references to objects that make up the content of the document.
- **Cross-reference table:** Contains a list of references to objects in the file to allow for random access and reuse of these objects.
- **Trailer:** Provides quick access to the cross-reference table and certain special objects, as PDFs should be read from the end.

There are two ways to update a PDF document. The first is to have the computer rewrite the whole file upon saving and overwriting the old document. The second is to use incremental updates, in which case the additions are added to the end of the file after the trailer, using additional body, cross-reference, and trailer sections for the update.

A PDF document is structured as a collection of objects. These objects are organized into a hierarchy and are interconnected to form the document. File page objects specifically pertain to the content and structure of individual pages within the PDF document. File page objects are important because they encapsulate all the details needed to render and interact with a particular page within the PDF document. They are integral to the rendering process and are essential for any operations involving manipulation, display, or analysis of the document [3] [[5]. File page objects consist of the following components:

- **Content Stream:** The content stream is a sequence of instructions describing how to display the page. It typically includes instructions for drawing text, images, and graphics on the page. We use the content stream in our proposed prototype to generate a Merke Tree (refer to section 3.2).
- **Resources:** These are objects that define the resources available to the page, such as fonts, images, and other reusable assets.
- **Media Box:** Defines the boundaries of the physical medium on which the page is to be printed.
- **Crop Box:** Defines the region to which the contents of the page should be clipped.
- **Rotation:** Specifies the rotation angle of the page.
- **Annotations:** Optional elements that allow for interaction with the user or define actions to be taken when the document is opened, or certain events occur.

3.2 Implementation Details

Our proposed prototype can be used in two different ways with a PDF document. Firstly, to protect a PDF document using our developed technique, the PDF must be run through the prototype so that a hash can be generated and inserted into the PDF document. Secondly, to assess a PDF document for forgery or tampering, the PDF must be run through the prototype, and the detection process will determine if changes have been made.

The prototype was implemented using the Python language. The hashlib and Merkly libraries are used to generate the hashes. These two main processes can be run independently of each other. To read the file page objects of the PDF, we make use of the Portable Document Format Read and Write (PDFRW) library. The PDFRW library gives us the ability to read and write PDF files. PDFRW was chosen because it offers lower-level access to PDF structures, which can be beneficial for custom manipulation tasks. Additionally, for certain operations, PDFRW can be faster due to its low-level access, especially when handling large PDFs or complex tasks. Figure 1 outlines the general process flow of the prototype.

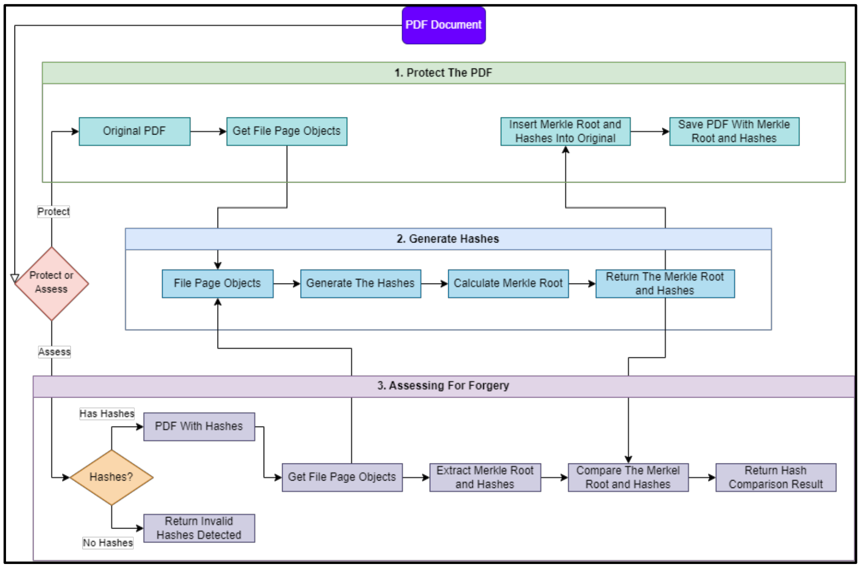


Fig. 1. Prototype Flow

The three main sub-components of the prototype will be discussed in more detail in the following sub-sections: in sub-section 3.2.1, we discuss how a PDF document is protected against tampering; in sub-section 3.2.2, we discuss how the hashes are generated for a PDF document; and in sub-section 3.2.3, we discuss how a PDF document is assessed for tampering.

3.2.1 Protect the PDF

The PDF document must have the calculated hashes to ensure the technique can detect and analyze tampering or forgery. The first phase, the "Protect The PDF" sub-component from Figure 1, is responsible for this process.

First, it will read the PDF document into the prototype using the PDFRW library, which produces a dictionary-like object for the parts of a PDF document. Once the PDF document has been read, we isolate the file page objects to use them for further processing. An iterative process of computing all the necessary hashes for a particular file page object will begin. For each page, the content stream of the file page object is then used to generate a Merkle tree. This is discussed in detail in section 3.2.3.

Following the calculation of the Merkle tree for the file page object content stream, a hash will be computed using the file page object itself. This is handled using the "2. Generate Hashes" sub-component showed in Figure 1, which is discussed in section

3.2.2. Before the hash can be computed for the file page object, it is first serialized and converted into a format that can be encoded into bytes. Serialization is the process of converting the file page objects and the root object of the PDF into a format that can be easily stored or transmitted, often as a sequence of bytes. Serialization is necessary to ensure that these objects are in a consistent and standard format that can be processed correctly when generating a hash. Some sub-objects of the file page object are excluded during serialization to prevent the calculation of a protected PDF from producing a false result, due to not all PDF document editors creating and updating a PDF document uniformly.

After the hash values are calculated, they are stored in an object structured as follows:

```
{‘object': hash value, 'root': hash value, 'leafs': list
of hash values}
```

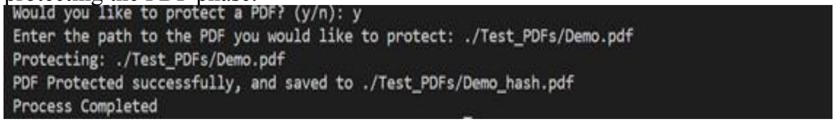
This object is discussed further in detail in section 3.2.2. The values associated with each heading in the object are the values that will be stored within the PDF document. This is done by creating three new keys in the relevant file page object and storing the relevant value inside its respective key. The keys created within the file page object are: ‘hashobject’, ‘hashroot’, and ‘hashleafs’. If there are four pages within the PDF document, this process will be repeated for all four pages.

The next step to protect the PDF involves generating a hash value for the root object, which is structured as follows:

```
{‘root': hash value, 'info': hash value}
```

This object will be discussed further in section 3.2.3. The values associated with each heading in the object are the values that will be stored within the PDF document in the root object. This is done by creating two new keys in the root object and storing the relevant value inside its respective key. The keys created within the file page object are: ‘hashroot’ and ‘hashinfo’.

Once the hashes have all been inserted into the relevant file page objects and the root object, we use the PDFRW library to save a new PDF document. The new PDF document is now the protected PDF document and can be used in the “3. Assessing for Forgery” sub-component. Figure 2 shows an example output of the prototype after the protecting the PDF phase.



```
Would you like to protect a PDF? (y/n): y
Enter the path to the PDF you would like to protect: ./Test_PDFs/Demo.pdf
Protecting: ./Test_PDFs/Demo.pdf
PDF Protected successfully, and saved to ./Test_PDFs/Demo_hash.pdf
Process Completed
```

Fig. 2. Outcome after successfully protecting a PDF

3.2.2 Generating the Hashes

The second phase, the “2. Generate Hashes” sub-component, is used in sub-components 1 and 3 (refer to Figure 1). It is used to generate the object that was discussed in sub-section 3.2.1. This object is used to protect and assess tampering within the PDF document. The elements of the object and how they are generated is discussed based on the order they appear within the object. For reference, the object looks like as follows: {'object': hash value, 'root': hash value, 'leafs': list of hash values}.

The 'object' hash value is calculated using the hashlib library's implementation of the SHA256 hashing algorithm. It is computed using the file page object after it has been serialized and encoded into a byte stream.

The 'root' and 'leafs' hash values are calculated simultaneously using the Merkly library's implementation of a Merkle root. This library utilizes the Keccak hashing algorithm. The Keccak hashing algorithm, also known as SHA-3 (Secure Hash Algorithm 3), is a cryptographic hash function designed to provide high security against various types of attacks, including collision, pre-image, and second pre-image attacks [16]. A Merkle root is the cumulative hash of all the hashes that comprise a Merkle tree. An example is seen in Figure 3. The 'root' hash value that we store within the PDF is the Merkle root of the Merkle tree is created using the content stream within the file page object.

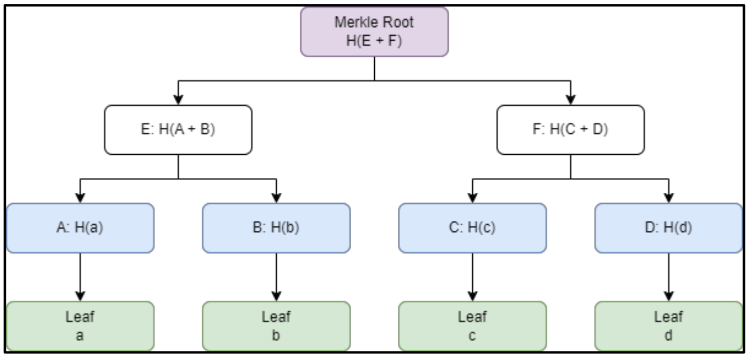


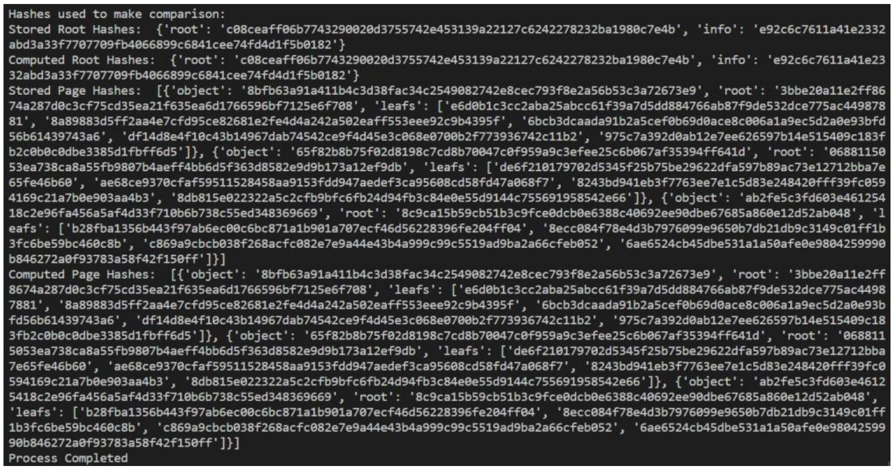
Fig. 3. Structure of the Merkle Tree

The content stream is divided into groups of 256 bytes, which are then used as the leaves within the Merkle tree. The 'leafs' hash values in the object are the previously calculated hash values of these groups. These values are stored within the PDF to localize where a change has been made within the file page object. If a change is made to the content of the file page object, the change can be localized to the nearest 256-byte group of the content stream.

Once all three groups of hash values have been calculated, they are returned to the sub-component that called the second phase, the “2. Generate Hashes” sub-component. This could be phase 1 or phase 3.

After the previously mentioned hash values have been stored in the PDF, the calculation of the hash values for the root and metadata will be done. This is done second to the hash values for the file page objects because those hashes are included when calculating the root hash value. The hash object that will be produced is:
`{‘root': hash value, 'info': hash value}`.

The root hash value is calculated by serializing the sub-objects of the Root and calculating the hash with the produced value using the SHA256 algorithm. The info hash value is calculated by serializing the metadata of a PDF document and again using the SHA256 algorithm. The hash object is returned so the values can be stored in the root object or used for comparison. Figure 4 shows the output of the generated hashes.



```
Hashes used to make comparison:
Stored Root Hashes:  {'root': 'c08ceaff06b7743290020d3755742e453139a22127c6242278232ba1980c7e4b', 'info': 'e92c6c7611a41e2332
abd3a33f7707709fb4066899c6841cee74fd4d1f5b0182'}
Computed Root Hashes:  {'root': 'c08ceaff06b7743290020d3755742e453139a22127c6242278232ba1980c7e4b', 'info': 'e92c6c7611a41e23
32abd3a33f7707709fb4066899c6841cee74fd4d1f5b0182'}
Stored Page Hashes:  [{'object': '8bfb63a91a411b4c3d38fac34c2549082742e8cec793f8e2a56b53c3a72673e9', 'root': '3bbe20a11e2ff86
74a287d0c3cf75cd35ea21f635ea6d1766596bf7125e6f708', 'leafs': ['e6d0b1c3cc2aba25abcc61f39a7d5dd884766ab87f9de532dce775ac449878
81', '8a89883d5ff2aa4e7cfd95ce82681e2fe4d4a242a502eaff553eee92c9b4395f', '6bcb3dcaada91b2a5cef0b69d0ace8c006a1a9ec5d2a0e93bfd
56b61439743a6', 'df14d8e4f10c43b14967dab74542ce9f4d45e3c068e0700b2f773936742c11b2', '975c7a392d0ab12e7ee626597b14e515409c183f
b2c0b0c0dbe3385d1fbff6d5']}, {'object': '65f82b8b75f02d8198c7cd8b70047c0f959a9c3efee25c6b067af35394ff641d', 'root': '06881150
53ea738ca8a55fb9807b4aeff4bb6d5f363d8582e9d9b173a12ef9db', 'leafs': ['de6f210179702d5345f25b75be29622dfa597b89ac73e12712bba7e
65fe46b60', 'ae68ce9370cfaf59511528458aa9153fdd947aedef3ca95608cd58fd47a068f7', '8243bd941eb3f7763ee7e1c5d83e248420fff39fc059
4169c21a7b0e903aa4b3', '8db815e022322a5c2cfb9bfc6fb24d94fb3c84e0e55d9144c755691958542e66']}, {'object': 'ab2fe5c3fd603e461254
18c2e96fa456a5af4d33f710b6b738c55ed348369669', 'root': '8c9ca15b59cb51b3c9fce0dcb0e6388c40692ee90dbe67685a860e12d52ab048', 'l
eafs': ['b28fba1356b443f97ab6ec00c6bc871a1b901a707ecf46d56228396fe204ff04', '8ecc084f78e4d3b7976099e9650b7db21db9c3149c01ff1b
3fc6be59bc460c8b', 'c869a9cbcb038f268acfc082e7e9a44e43b4a999c99c5519ad9ba2a66cfeb052', '6ae6524cb45dbe531a1a50afe0e9804259990
b846272a0f93783a58f42f150ff']}]
Computed Page Hashes:  [{'object': '8bfb63a91a411b4c3d38fac34c2549082742e8cec793f8e2a56b53c3a72673e9', 'root': '3bbe20a11e2ff
8674a287d0c3cf75cd35ea21f635ea6d1766596bf7125e6f708', 'leafs': ['e6d0b1c3cc2aba25abcc61f39a7d5dd884766ab87f9de532dce775ac4498
7881', '8a89883d5ff2aa4e7cfd95ce82681e2fe4d4a242a502eaff553eee92c9b4395f', '6bcb3dcaada91b2a5cef0b69d0ace8c006a1a9ec5d2a0e93b
fd56b61439743a6', 'df14d8e4f10c43b14967dab74542ce9f4d45e3c068e0700b2f773936742c11b2', '975c7a392d0ab12e7ee626597b14e515409c18
3fb2c0b0c0dbe3385d1fbff6d5']}, {'object': '65f82b8b75f02d8198c7cd8b70047c0f959a9c3efee25c6b067af35394ff641d', 'root': '068811
5053ea738ca8a55fb9807b4aeff4bb6d5f363d8582e9d9b173a12ef9db', 'leafs': ['de6f210179702d5345f25b75be29622dfa597b89ac73e12712bba
7e65fe46b60', 'ae68ce9370cfaf59511528458aa9153fdd947aedef3ca95608cd58fd47a068f7', '8243bd941eb3f7763ee7e1c5d83e248420fff39fc0
594169c21a7b0e903aa4b3', '8db815e022322a5c2cfb9bfc6fb24d94fb3c84e0e55d9144c755691958542e66']}, {'object': 'ab2fe5c3fd603e4612
5418c2e96fa456a5af4d33f710b6b738c55ed348369669', 'root': '8c9ca15b59cb51b3c9fce0dcb0e6388c40692ee90dbe67685a860e12d52ab048',
'leafs': ['b28fba1356b443f97ab6ec00c6bc871a1b901a707ecf46d56228396fe204ff04', '8ecc084f78e4d3b7976099e9650b7db21db9c3149c01ff
1b3fc6be59bc460c8b', 'c869a9cbcb038f268acfc082e7e9a44e43b4a999c99c5519ad9ba2a66cfeb052', '6ae6524cb45dbe531a1a50afe0e98042599
90b846272a0f93783a58f42f150ff']}]
Process Completed
```

Fig. 4. Output of Generated Hashes

The following sub-section discusses the process of assessing a PDF for tampering or forgery.

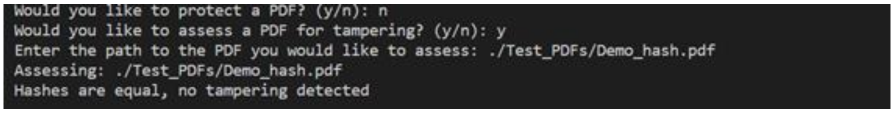
3.2.3 Assessing for PDF forgery

The second primary function that the prototype can accomplish is the assessment of tampering within a PDF document. For phase 3, “3. Assessing for Forgery”, to successfully detect tampering, it requires that phase 1 be run on the PDF document before the suspected tampering was done.

To perform the assessment, the prototype will read the PDF using the PDFRW library and extract the file page and root objects. It will then extract the hash values stored in the root object associated with the following keys: ‘hashroot’ and ‘hashinfo’. Following this, it will extract all the groups of hashes for each file page object in the PDF document from the relevant keys. The keys that will be extracted are: ‘hashobject’, ‘hashroot’, and ‘hashleafs’.

After extracting the hash values from the root object and the file page objects, the keys will be removed from the respective objects before calculating the hash value for the relevant object. The calculated hash values will be compared to the previously extracted hash values to determine if the file has been tampered with.

A relevant message indicating the presence or absence of tampering will be displayed. Should tampering be detected in the PDF document, specifically within the root object, or the metadata of the PDF document, a message indicating this will be displayed. If tampering has been detected, the page number where the changes were made and in which group of the byte stream the changes will be displayed. Figure 5 shows the output of the assessment for forgery.



```
Would you like to protect a PDF? (y/n): n
Would you like to assess a PDF for tampering? (y/n): y
Enter the path to the PDF you would like to assess: ./Test_PDFs/Demo_hash.pdf
Assessing: ./Test_PDFs/Demo_hash.pdf
Hashes are equal, no tampering detected
```

Fig. 5. Output when assessing for forgery

4 Experimentation and Results

This section focuses on evaluating the prototype's effectiveness, which is crucial for determining its practicality and validity. It outlines the structure, starting with an explanation of changes made to PDFs for testing purposes, covering alterations to text, images, and metadata.

4.1 Results

The tests include scenarios such as text addition, alteration, and removal, image insertion, and manipulation, as well as metadata changes [14]. Each type of alteration is tested individually and then combined in a comprehensive test file. This thorough evaluation aims to assess the prototype's ability to detect various alterations accurately.

Tables 1 to 4 below tabulate the results of all the tests conducted using the prototype. All PDF documents used in the test cases described below were first protected by the prototype, and then the alterations were made. The only exception to this protection is the first test, which explicitly tests for the absence of the hashes. The PDF documents were created using Microsoft Word and Adobe Acrobat. PDF names with an asterisk (*) at the beginning denote that the PDF was created with Microsoft Word; those without were created with Adobe Acrobat. All changes to the text, images, and metadata within a PDF were done using Adobe Acrobat. The file page objects will change depending on the changes made to the text and images within a PDF document.

Table 1. Experimentation Results: Not Protected

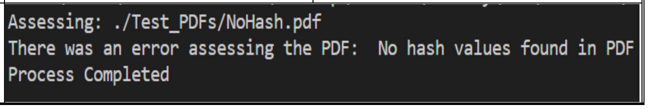
Test Name	PDF Name	Test Description
No Hashes	NoHash hash.pdf	This test assesses the ability of the prototype to detect the absence of hashes in a PDF.
Result	 Assessing: ./Test_PDFs/NoHash.pdf There was an error assessing the PDF: No hash values found in PDF Process Completed	

Table 1 shows that the prototype can detect when a PDF does not have the hashes stored within the file. It also indicates that it cannot assess such a file because no hashes are stored.

Table 2. Experimentation Results: Text Alteration

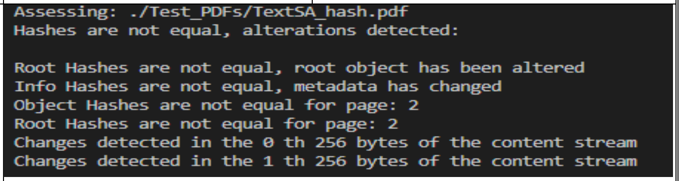
Test Name	PDF Name	Test Description
Single addition	* TextSA hash.pdf	This test sees the addition of a singular line of text on the 2nd page of the PDF document.
Result	 Assessing: ./Test_PDFs/TextSA_hash.pdf Hashes are not equal, alterations detected: Root Hashes are not equal, root object has been altered Info Hashes are not equal, metadata has changed Object Hashes are not equal for page: 2 Root Hashes are not equal for page: 2 Changes detected in the 0 th 256 bytes of the content stream Changes detected in the 1 th 256 bytes of the content stream	
Multiple addition	* TextMA hash.pdf	This test sees the addition of multiple singular lines of text on the 2nd page of the PDF document.

Table 2 shows that the prototype can detect when the text in a PDF has been altered. Be this an addition, update, deletion, or a combination of all of the above. The messages displayed indicate which page or pages the text has been altered on and which parts of the content stream have been altered.

Table 3. Experimentation Results: Image Alteration

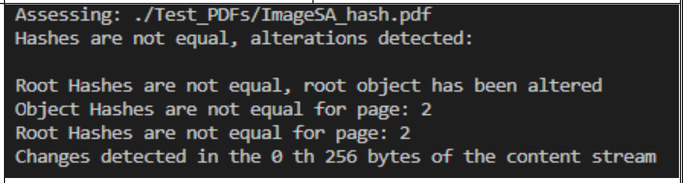
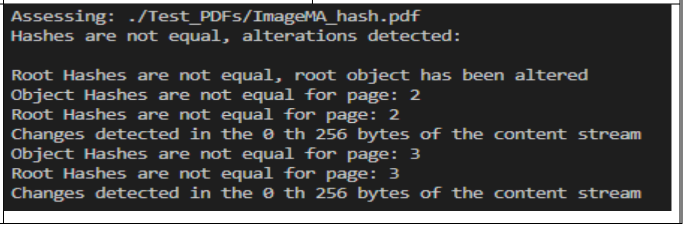
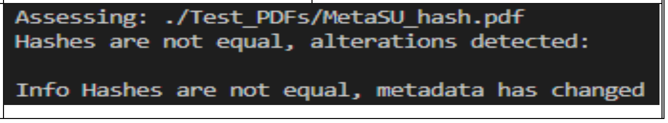
Test Name	PDF Name	Test Description
Single addition	ImageSA hash.pdf	This test sees the addition of a singular image on the 2nd page of the PDF document.
Result	 Assessing: ./Test_PDFs/ImageSA_hash.pdf Hashes are not equal, alterations detected: Root Hashes are not equal, root object has been altered Object Hashes are not equal for page: 2 Root Hashes are not equal for page: 2 Changes detected in the 0 th 256 bytes of the content stream	
Multiple addition	ImageMA hash.pdf	This test sees the addition of multiple images on the 2nd and 3rd page of the PDF document.
Result	 Assessing: ./Test_PDFs/ImageMA_hash.pdf Hashes are not equal, alterations detected: Root Hashes are not equal, root object has been altered Object Hashes are not equal for page: 2 Root Hashes are not equal for page: 2 Changes detected in the 0 th 256 bytes of the content stream Object Hashes are not equal for page: 3 Root Hashes are not equal for page: 3 Changes detected in the 0 th 256 bytes of the content stream	

Table 3 shows results when the prototype successfully detects when an image or images in a PDF have been altered, whether it is an addition, update, deletion, or a combination of all of the above. The messages displayed indicate which page or pages the text has been altered on and which parts of the content stream have been altered.

Table 4. Experimentation Results: Metadata Alteration

Test Name	PDF Name	Test Description
Single Update	* MetaSU hash.pdf	This test sees the addition of a singular metadata element in a PDF document.
Result	 Assessing: ./Test_PDFs/MetaSU_hash.pdf Hashes are not equal, alterations detected: Info Hashes are not equal, metadata has changed	

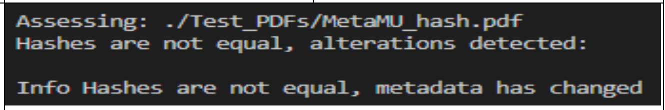
Multiple Update	* MetaMU hash.pdf	This test sees the update of multiple metadata elements in the PDF document.
Result	 Assessing: ./Test_PDFs/MetaMU_hash.pdf Hashes are not equal, alterations detected: Info Hashes are not equal, metadata has changed	

Table 4 shows that the prototype can detect tampering with the metadata of a PDF document. The message displayed indicates that the metadata has been tampered with.

4.2 Evaluation

This section provides a critical evaluation of the prototype and its ability to detect tampering within a PDF document. First, we discuss the general ability of the prototype. Then we discuss the validity of the prototype's results and its limitations.

4.2.1 General Analysis of the Prototype.

The prototype successfully detected changes in the three main categories we tested: changes to text, changes to images, and changes to metadata. These changes were all made using Adobe Acrobat. This means that we can only confirm that the prototype successfully detects changes when they are made using Adobe Acrobat. Different PDF editors produce PDF files with different layouts of underlying objects. However, the prototype should be able to detect the aforementioned changes regardless of the PDF editor, because the protected PDF documents are produced using the PDFRW library.

As seen in Section 3, in Tables 1 to 4, the output results describe on which page the change was detected or, in the case of metadata, that a change in the metadata object has been made. Further expansion on the indication of which page the change is on also indicates which parts of the content stream have been altered.

One thing to note is how the PDFRW library writes a PDF and the order of the objects in the PDF document differ from other PDF editors. For this reason, the file page objects, and the root object must be serialized before they can be used to produce a hash. In the current iteration of the prototype, specific sub-elements are selected and serialized for use in creating the hash for the object. It is worth noting that our proposed method works in instances when the PDF is protected (refer to Section 3.2). If it is not protected, then it would not be possible to detect where alterations were made.

4.2.2 Evaluation of Prototype Validity.

Since the objects must be serialized before they are used for the generation of the hash, any addition or alteration that does not affect the selected elements will not be detected by the prototype. Because all PDF editors save the PDF with a different underlying structure, we cannot serialize the entire object as-is because the order of these objects will impact the produced hash. This same issue with adding custom elements to a PDF can be applied to the PDF metadata, where one can add their custom metadata elements. The issue with not being able to consistently create a hash using the object is a complex task that requires extensive research into the most consistent way to convert it into a uniform input for the hash function.

When protecting the PDF document, a new file, a replica of the original that includes the hashes embedded into the file page objects, is created. This in itself is a form of tampering with the PDF document. There is no feasible way to insert the hashes into the original file without tampering. For the sake of later assessment by a document tampering specialist, making a copy of the original file was chosen as the route. This technique of protecting the PDF relies on the fact that only the protected PDF will be the one tampered with by a malicious individual.

We rely on the content stream for the majority of the detection of changes to the content of a particular PDF document. This will only detect a change to the physical text or the visible image to the user. If you were to change the font for a particular line or the entire document, then the prototype would not be able to detect this. This stems from the issue mentioned in the previous paragraph, where not all file page object elements can be used to generate the object's hash. While the content stream is part of the elements that make up a file page object, the current prototype uses its parent element when creating the hash for that particular file page object.

The prototype succeeds at detecting tampering with a PDF document. However, this is limited to text, images, and some types of alterations to metadata and file page objects. There are still many ways to modify the PDF document that the prototype would not detect, such as adding JavaScript to the PDF document.

5 Summary and Conclusion

The paper presented a technique to detect tampering or forgery in PDF documents using file page objects. The development of a prototype for this purpose is highlighted as a significant contribution. The developed prototype successfully utilized these file page objects to detect tampering within a PDF document. The file page objects were employed to generate hashes, which can be used to detect and analyze the presence of tampering within a PDF document.

Future research suggestions include generalizing the prototype to accommodate various PDF structures, investigating additional types of changes such as JavaScript additions and signature alterations, and considering the protection of scanned PDF documents. These potential avenues aim to enhance the effectiveness and scope of tampering detection within PDF files.

References

1. G. Saju and K. Sreenimol, “An effective method for detection and localization of tampering,” International Journal of Information, vol. 8, pp. 152–154, no. 2, 2019.
2. P. Johnston and E. Elyan, “A review of digital video tampering: From simple editing to full synthesis,” Digital Investigation, vol. 29, pp. 67–81, 2019.
3. T. Bienz, R. Cohn, and C. Adobe Systems: Mountain View, Portable document format reference manual. Citeseer, 1993.
4. N. A. A. Usop, S. I. Hisham, and J. M. Zain, “An implementing of zigzag pattern in numbering watermarking bits for high detection accuracy of tampers in document,” Indian Journal of Computer Science and Engineering (IJCSE), vol. 13, pp. 1733-1751, 2022.
5. U. Khadam, M. M. Iqbal, M. A. Habib, and K. Han, “A watermarking technique based on file page objects for pdf,” in 2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM), pp. 1–5, IEEE, 2019.
6. I. B. Senkyire and Q.-A. Kester, “A cryptographic tamper detection approach for storage and preservation of forensic digital data based on sha-384 hash function,” in 2021 International Conference on Computing, Computational Modelling and Applications (ICCMA), pp. 159–164, IEEE, 2021.
7. D.S. Popescu, “Hiding malicious content in pdf documents,” Journal of Mobile, Embedded and Distributed Systems, 3 (3) (2011), pp. 120-127, 2012.
8. V. Mladenov, C. Mainka, K. Meyer zu Selhausen, M. Grothe, and J. Schwenk, “1 trillion-dollar refund: How to spoof pdf signatures,” in Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, pp. 1–14, 2019.
9. P. Dikanev and Y. Vybornova, “Method for protection of pdf documents against counterfeiting using semi-fragile watermarking,” in 2021 International Conference on Information Technology and Nanotechnology (ITNT), pp. 1–4, IEEE, 2021.
10. C. Mainka, V. Mladenov, and S. Rohlmann, “Shadow attacks: Hiding and replacing content in signed pdfs.,” in 28th Annual Network and Distributed System Security Symposium, NDSS 2021, The Internet Society, 2021.
11. C. Gao, X. Wan, C. Guo, B. Wu. "Blockchain-based PDF File Copyright Protection and Tracing." Springer Handbook of peer-to-peer networking, 2023.
12. N. Josue, F. Tchakounte, PM. Buhendwa, M. Atemkeng. "Fals-Ism: A Graph Isomorphism Framework for Multi-Level Detection of Falsified PDF Documents." Journal of Computer Science 19.5: pp. 667-676, 2023.
13. Z. Jiang, H. Wang, SY. Han. “A robust PDF watermarking scheme with versatility and compatibility.” Multimedia Tools and Applications, pp. 1-27, 2024.
14. AH. Kamakshi, and S. Grandhi. "Text classification from PDF documents." International Research Journal of Modernization in Engineering Technology and Science 3: pp. 58-63, 2021.
15. Bitar, Ahmad W., et al. "Blind digital watermarking in PDF documents using Spread Transform Dither Modulation." Multimedia Tools and Applications 76: pp. 143-161, 2017.

16. A. Gholipour, & S, Mirzakuchaki. “High-speed implementation of the Keccak hash function on FPGA”. International Journal of Advanced Computer Science, 2(8), pp: 303-307, 2012

Machine Learning-based NLP for Emotion Classification on a Cholera X Dataset

Paul Jideani[1,2][0000-0001-5836-6660] and Aurona Gerber[1,3][0000-0003-1743-8167]

[1] Department of Computer Science, University of the Western Cape, South Africa
[2] Boston City Campus, Stellenbosch, South Africa
[3] Center for AI Research (CAIR), South Africa
pcijideani@gmail.com

Abstract. Recent social media posts on the cholera outbreak in Hammanskraal have highlighted the diverse range of emotions people experienced in response to such an event. The extent of people's opinions varies greatly depending on their level of knowledge and information about the disease. The documented research about Cholera lacks investigations into the classification of emotions. This study aims to analyze the emotions conveyed in social media posts regarding Cholera. A dataset of 23,000 posts was extracted and pre-processed. The VADER sentiment analyzer library was applied to determine the emotional significance of each text. Additionally, Machine Learning (ML) models were applied for emotion classification, including Long short-term memory (LSTM), Logistic regression, Decision trees, and the Bidirectional Encoder Representations from Tranformers (BERT) model. The results of this study demonstrated that LSTM achieved the highest accuracy of 76%. Emotion classification presents a promising tool for gaining a deeper understanding of the impact of Cholera on society. The findings of this study might contribute to the development of effective interventions in public health strategies.

Keywords: Cholera, machine learning, emotion classification, natural language processing.

1 Introduction

Cholera is a waterborne infectious disease that causes severe diarrhoea and vomiting in humans [1, 2]. It remains a persistent menace in many parts of Africa and Asia. Cholera is transmitted by drinking water or eating food contaminated by bacteria [3]. According to the WHO, the disease causes 1.3 to 4 million cases and 21,000 to 143,000 deaths worldwide yearly, mostly in underdeveloped nations, including India (2007), Iraq (2008), Congo (2008), and Zimbabwe (2008-2009) [1, 2, 4]. One major concern regarding Cholera is its potential to cause large-scale outbreaks that spread rapidly [5, 6]. It has been a source of concern for public health practitioners as outbreaks can overwhelm healthcare systems and strain available resources [7]. The recent cholera outbreak in Hammanskraal, South Africa, has underlined the importance of studying how the general public perceives and responds to the disease. According to the Department of Health, the outbreak started in May 2023, hospitalized 95 people and had a death toll of 31 as of 8 June 2023 [4, 8, 9]. While most research has focused on discovering treatments and vaccinations for the disease, there is growing interest in studying public views and emotions surrounding Cholera, which is the focus of this study [10, 11]. This study compared the results of ML-based NLP techniques to analyze public sentiment, specifically the expressed emotions in X (previously Twitter) data on Cholera.

The rest of the paper is organized as follows: Section 2 provides a summary of the relevant literature. Section 3 covers materials and methods, Section 4 outlines the results and analysis, and Section 5 draws conclusions.

2 Literature Review

Public opinion mining, or sentiment analysis, extracts and analyses subjective information from various sources to understand and evaluate public sentiment, attitudes, and opinions towards a particular topic, product, service, or event [12, 13]. It involves the use of natural language processing (NLP) and Machine Learning (ML) techniques to analyze large volumes of textual data, such as social media posts, online reviews, news articles, and customer feedback [14]. Public opinion mining aims to understand people's perceptions and feelings about a specific topic, whilst Sentiment analysis specifically has been used to analyze and classify the sentiment expressed in texts, allowing organizations, businesses, and governments to understand public perception, monitor brand reputation, assess customer satisfaction, predict trends, and make data-driven decisions [15].

Public opinion mining on social media has become increasingly important in today's digital age. Social media platforms have transformed the way people communicate and express their opinions and experiences [16, 17] and organizations use this source of user-generated content to extract insights about public sentiments, attitudes, and opinions, often in real-time [18, 19]. These insights are invaluable for making data-driven decisions, refining marketing strategies, staying competitive in the market and even providing input to policymakers and government entities [20–23]. Public opinion mining on social media has implications beyond marketing and business. These insights can guide decision-making processes, help shape policies that align with public expectations, and contribute to better governance [24, 25].

Several studies have conducted sentiment analysis on COVID-19 and other disease datasets. Jatla and Avula [26] used a Hybrid Deep Sentiment Analysis (HDSA) model to analyze the sentiments in COVID-19-related tweets. The model was trained on a dataset of COVID-19 tweets, achieving a classification accuracy of 94%. Hossain et al. [27] developed a DL-based technique for analyzing COVID-19 tweets, using Bidirectional Gated Recurrent Unit (BiGRU). The model was trained on an improved dataset, and it achieved an accuracy of 87%. Singh et al. [28] conducted a study COVID-19 Twitter data using sentiment analysis and ML techniques to help predict outbreaks and epidemics. The paper discusses various ML techniques such as ME, DT, Support Vector Machine (SVM), and Naive Bayes for sentiment analysis. The article, however, does not provide any empirical evidence or case studies to support the effectiveness of the proposed approach. Kaur et al. [29] proposed a sentiment analysis DL algorithm for classifying tweets into positive, negative, and neutral sentiment scores using the Hybrid Heterogeneous Support Vector Machine (H-SVM) method.

Ronchieri et al. [30] employed Natural Language Processing (NLP), sentiment analysis, and topic modelling techniques to analyze a dataset comprising 369,472 tweets. The findings revealed prevalent emotions such as fear, trust, and disgust, while prominent discussions centered around topics like malaria, influenza, and tuberculosis. The analytical methods applied included the Latent Dirichlet Allocation (LDA) model and TF-IDF vectorization.

Han & Thakur [31] performed sentiment analysis using a methodology that involved analysing 12,028 tweets focusing on the Omicron variant. The results indicated that 50.5% of the tweets conveyed a neutral sentiment, and the predominant language used was English, accounting for 65.9%. Oladipo et al. [32] presented a study wherein sentiment analysis was conducted using the NRC lexicon approach. The results revealed an overall positive sentiment towards the lockdown exercise, based on an analysis of 22,249 tweets sourced from national stakeholders and the general public. Shaalan et al. [33] presented a sentiment analysis study focused on evaluating sentiments expressed in Arabic tweets related to COVID-19. The study employed various models, starting with data acquisition followed by text pre-processing. Term Frequency Inverse Document Frequency (TF-IDF) was utilized to generate feature vectors. Multiple classifiers, including Naïve Bayes, Support Vector Machine, Logic Regression, Random Forest, and K-Nearest Neighbour, were compared through experiments. Performance evaluation was conducted using metrics such as Precision, Accuracy, Recall, and F1 Score. The most effective model achieved an accuracy of approximately 84%.

As is indicated in the summary of related work above, the use of NLP and ML for opinion mining to detect emotions in X (previously Twitter) data is a well-recognized technique, and this study adopted this approach to detect emotions regarding the Cholera outbreak in South Africa.

3 Materials and Methods

Figure 1 illustrates the research methodology framework, encompassing dataset management and the implementation of ML algorithms. Dataset management consists of dataset collection, pre-processing, and balancing. These stages are fundamental in ensuring the quality, consistency, and unbiased representation of the data. The application of ML algorithms presents the specific algorithms employed: Bidirectional Encoder Representations from Transformers (BERT), Long Short-Term Memory (LSTM), Logistic Regression, and Decision Tree.

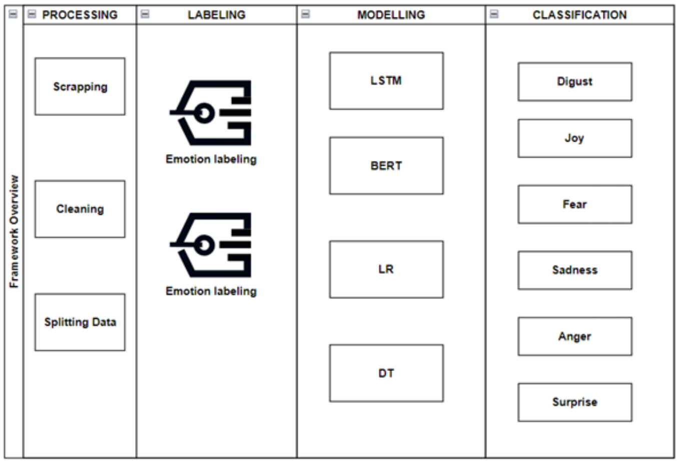


Fig. 1. Research Methodology Overview

- **Dataset Collection:** The Twitter API was used to extract tweets using #Hammanskraal # CholeraOutbreak, #water, #Cholera hashtags during a specific time range using the Tweepy library. The information extracted from each returned tweet included the tweet ID, text, and creation date. Between April 2023 and July 2023, a total of 23000 tweets were collected. The script included monitoring statements to print the current date and the total number of collected tweets. After completing the scraping, a Pandas DataFrame was generated from the list of tweets, which was then saved to a CSV file.
- **Dataset Pre-processing:** The collected dataset underwent preprocessing for emotion analysis, starting with training a tokenizer on the entire dataset to learn vocabulary and word-to-index mappings. This trained tokenizer was then used to convert both the training and test data into sequences of numerical indices, representing the words. Padding sequences ensured uniform lengths for input consistency in neural network models. Additionally, labels for both sets were transformed into dummy variables using one-hot encoding. The dataset was further cleaned by removing duplication, extending contractions, eliminating stopwords, stripping numeric values and punctuation, removing extra spaces, and deleting greetings and other unnecessary words.
- **Emotion Labelling:** The pysentimiento library was used to analyze emotion within tweets. Utilizing Ekman's basic wheel of emotions, the objective was to discern six specific emotions within a corpus of tweets. The emotion with the highest probability was extracted from the predicted emotions, excluding the "others" category, and appended to the predicted emotions list. Throughout the iteration, the output was cleared and the current number of predicted emotions was printed to monitor the progress. Once all tweets were analyzed, the predicted emotions the results were saved as a CSV file. Algorithm 1 illustrates the emotion labelling algorithm.

Algorithm 1 LABELING STEPS

```
Start:
    # Iterate through each row of the DataFrame
    for each row in emo_analysis_data1:
        # Predict emotion using the emotion analyzer on the 'Text' column
        output = emotion_analyzer.predict(row['Text'])

        # Extract the emotion with the highest probability, excluding
"others"
        emotion = [emotion for emotion in output.probas.keys() if emotion
!= "others"]
        highest_emotion = max(emotion, key=lambda emotion: output.pro-
bas[emotion])

        # Append the highest emotion to the predicted_emotions list
        predicted_emotions.append(highest_emotion)

    # End emotion analysis process
    End:

    # Add the predicted_emotions list as a new column called 'Emotion' to
the DataFrame
    emo_analysis_data1['Emotion'] = predicted_emotions
```

Disgust, Joy, Fear, Sadness, Anger, and Surprise were the types of emotions detected in our dataset. As shown in Figure 4, Disgust was the most prevalent emotion with 38.8% of all emotions available, while Surprise was the least expressed emotion with 1.8%.

Dataset Balancing: The data was balanced for emotion analysis using a two-step approach. Initially, the dataset was split into training and test sets, followed by identifying both majority and minority classes within the training set based on the target variable. The minority class then underwent an oversampling process to address class imbalance, achieved by generating additional instances of the minority class. This balanced representation helped prevent bias towards the majority class and enhanced the machine learning model's predictive accuracy. The oversampling was implemented using the resample() function from the sci-kit-learn library, duplicating instances from the minority class until its size matched that of the majority class. Subsequently, the upsampled minority class was merged with the original majority class, resulting in a more balanced training dataset. Finally, the effectiveness of the oversampling technique was confirmed by analyzing the new distribution of class counts.

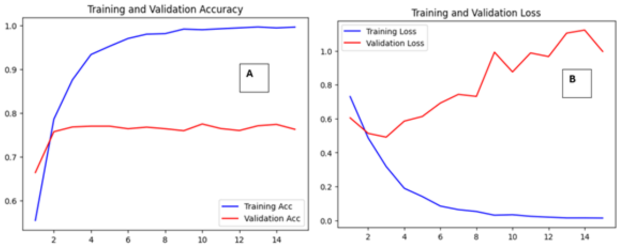


Fig. 2. Training and Validation Accuracy

Figures 2 A & B provide a line plot using the Matplotlib library to visualize the training and validation accuracy across various epochs during the training of a machine learning model. The x-axis denotes the epochs, which represent the iterations of the training process. On the y-axis, the accuracy values are depicted. The training accuracy is depicted in blue, while the validation accuracy is shown in red. The legend helps distinguish between the two lines, specifying which corresponds to the training accuracy and which to the validation accuracy.

The categorical_crossentropy loss function was chosen to evaluate the disparity between predicted and actual labels, while the Adam optimizer was employed to refine the model's parameters during training, with the objective of minimizing loss and improving prediction accuracy. The training procedure for the emotion analysis model included the instantiation of a History object to track training progress, along with the definition of a batch_size variable to determine mini-batch size. A "class_weights" dictionary was also established to address class imbalance, assigning weights to emotion classes and prioritizing the minority class during training. Finally, the model was trained using the "fit()" function, incorporating the defined batch size and class weights.

The performance of the emotion analysis model was evaluated through several methods. Initially, predictions for the test set labels were generated using the model's "predict()" function, and these predictions were structured into a dataframe for analysis. A confusion matrix was then created to evaluate the model's classification accuracy, providing details on true positives, true negatives, false positives, and false negatives. Furthermore, a classification report was generated to present precision, recall, F1-score, and support values for each emotion class, offering a comprehensive assessment of the model's performance. These evaluation techniques provided valuable insights into the model's strengths and areas for improvement in accurately classifying emotion labels.

3.1 Machine Learning Algorithms

This study analyzed four DL models: LSTM, Logistic Regression, Decision Tree, and BERT. A brief description of the four algorithms follows.

1. **Long Short-Term Memory (LSTM):** LSTMs address the vanishing gradient problem in traditional RNNs by introducing gating mechanisms. These mechanisms allow LSTMs to selectively retain and forget information over long periods [34]. As a result, LSTMs are better at capturing

long-term dependencies, which is essential for many NLP tasks, such as language modelling and machine translation [35]. The LSTM (Long Short-Term Memory) model architecture used in this study consists of three LSTM layers and two fully connected layers. Each LSTM layer is configured with an embedding layer having a dimension of 64, an input length of 29, and a vocabulary size of 100,000.

2. **Bi-directional Encoder Representations from Transformers (BERT):** BERT is a pre-trained transformer-based language model that captures bi-directional context from text. It is trained on massive text data and learns to generate deep contextualized word embeddings [36]. These contextualized embeddings have been proven to be highly effective in various downstream NLP tasks. BERT has achieved remarkable results in tasks like question-answering, natural language inference, and named entity recognition [37]. The key idea behind BERT is bi-directionality, which allows the model to consider the entire context of a word by looking both to the left and right of it. Unlike traditional language models that process text in a unidirectional manner, BERT is pre-trained using a masked language modelling objective [38].
3. **Logistic Regression (LR):** The Logistic Regression model is a simple yet powerful linear model that is widely used in binary classification tasks [39]. It is a probabilistic model that predicts the probability of an instance belonging to a specific class [40]. Logistic Regression can efficiently handle feature interactions and provide interpretable results, making it a popular choice for emotion analysis.
4. **Decision Tree:** Decision Trees are a non-parametric supervised learning method that learns a hierarchical structure of if-else rules to make predictions [41]. Decision Trees can handle both categorical and numerical features and capture complex feature interactions. They can also provide interpretable rules, making them suitable for emotion analysis tasks where it is important to understand the reasoning behind the predictions [42, 43].

3.2 Experimental Setup and Performance Metrics

Various experiments were conducted to assess the effectiveness of the four models developed in this study. The models were implemented using the Keras library and Python. The cleaned dataset comprised 19,077 tweets, which were split into 80% for training (15,262 instances) and 20% for testing (3,815 instances). The models were trained on the training set and evaluated on the testing set.

Table 2 training parameters used for each model

Model	Training Parameters

LSTM	embed_dim = 128
	lstm_out = 192

	max_fatures = 100000
	dropout=0.4
	Dense(2,activation='softmax'))
	batch_size = 128

Logistic Regression	cv=5 Number of iteraations = 1000

Decision Trees	criterion='gini'
	splitter=’best’
	min_samples_split=2, min_samples_leaf=1,
	Dropout rate = 0.5

BERT	Batch_size = 64
	Epochs = 5
	Dropout rate = 0.2
	Number of dense layers = 32

Table 2 overviews the training parameters used for various machine learning models, including LSTM, Logistic Regression, Decision Trees, and BERT. Specific parameters such as embedding dimension, LSTM output dimension, maximum features, dropout rate, batch size, cross-validation folds, number of iterations, criterion for decision trees, and batch size for BERT are specified for each model. These parameters are crucial in determining the model's performance and effectiveness in learning from the training data and making predictions.

The assessment of the model performance was achieved through the utilization of performance evaluation metrics. Numerous metrics have been introduced in documented research, each focusing on specific facets of algorithmic performance. Hence, for every machine learning problem, a suitable set of metrics is essential for accurate performance evaluation. In this study, we employ several standard metrics commonly used for classification problems to derive valuable insights into algorithm performance and facilitate a comparative analysis. This study adopted four performance metrics: accuracy, precision, recall, and F1 score. The metrics can be calculated using equations (1) - (4).

$$Precision = \frac{TP}{TP + FP} (1) \qquad Recall = \frac{TP}{TP + FN} (2)$$

$$F1 - score = 2 * \frac{Precsion * Recall}{Precision + Recall} (3) \qquad Accuracy = \frac{TP + TN}{TP + TN + FP + FN} (4)$$

4 Results and Analysis

Experiments were conducted to assess the effectiveness of LSTM, DT, LR and BERT. Additionally, a comparative analysis is provided to highlight the relative performance of these four models.

Figure 3 and Figure 4 depict the resulting distribution of emotions based on Ekman's Basic Wheel in the X cholera datasets. Notably, 691 instances conveyed 'anger,' reflecting discontent, possibly because the authorities handled the Cholera crisis. A substantial number of tweets, totalling 5822, expressed 'disgust' towards the unfolding cholera outbreak. 'Fear' was evident in 3172 tweets, signifying anxiety about the impact of the Cholera outbreak. On the positive spectrum, 4192 tweets suggested potential optimistic developments or an improving situation. The emotion of 'sadness' emerged in 853 tweets reflecting the adverse effects of the cholera outbreak on individuals. Lastly, 270 tweets conveyed 'surprise,' at the occurrence of a cholera outbreak in the modern age.

4.1 Performance Analysis

Table 1 illustrates that LSTM achieved a classification accuracy of 76%, indicating its ability to correctly predict class labels for a substantial portion of the dataset. Additionally, LSTM exhibited a precision, recall, and F1-score of 75%, 81%, and 78%, respectively, highlighting its high overall accuracy and balance between precision and recall. The precision of 75% suggests that 75% of instances predicted as positive were true positives, while the recall of 81% indicates successful capture of actual positive instances. The F1-score of 78% further confirms LSTM's robust performance by considering false positives and false negatives, showcasing its effectiveness in classification with minimal errors.

Results indicate that LR produced a classification accuracy of 60%. LR correctly classified 60% of the emotions in the dataset. The precision of 59% signifies that, among the instances predicted as positive, 59% were indeed true positives. On the other hand, the recall of 86% suggests that LR successfully captured a substantial proportion of actual positive instances. The F1-score of 70%, the harmonic mean of precision and recall, provides a balanced assessment by considering false positives and false negatives. While LR demonstrated high recall, indicating its ability to identify positive instances, the precision was relatively lower, implying a higher rate of false positives. This underlines LR's potential for improving the trade-off between precision and recall.

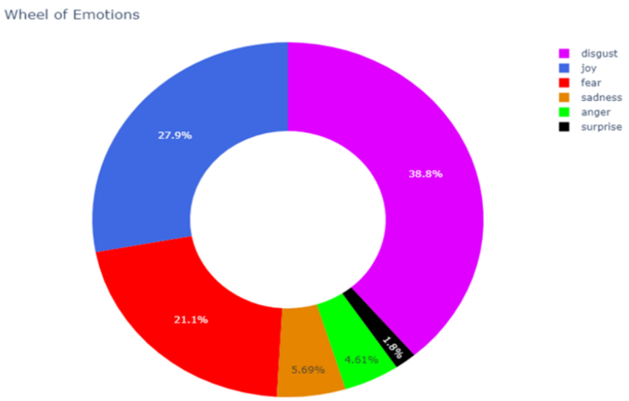


Fig. 3. Wheel of emotions expressed from the dataset

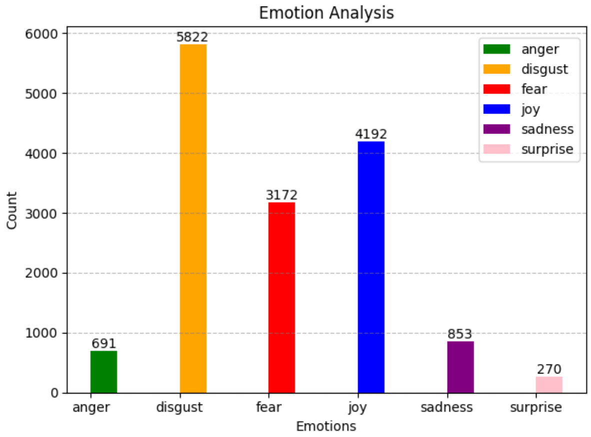


Fig. 4. Visualization of the number of emotions

The results show that the Decision Tree (DT) produced a classification accuracy of 56%. Examining precision, recall, and F1-score offers deeper insights into the model's performance beyond accuracy. The precision of 59% indicates that among instances predicted as positive by DT, 59% were true positives. Simultaneously, the recall of 86% suggests that DT effectively captured a substantial proportion of actual positive instances. The F1-score, which combines precision and recall, stands at 70%, offering a balanced evaluation. Notably, the performance metrics for DT closely resemble those of Logistic Regression, indicating a comparable ability to identify positive instances. However, similar to LR, the precision-recall trade-off should be considered for potential refinements in DT's performance.

BERT achieved a classification accuracy of 66%, showcasing its fundamental capability in predicting class labels. However, a more comprehensive assessment includes precision, recall, and F1-score. With a precision of 65%, BERT accurately identified 65% of positive instances among those predicted as positive. Furthermore, its recall of 68% indicates effective capture of actual positive instances. The balanced F1-score of 67% consolidates BERT's performance evaluation. These results underline BERT's effectiveness in emotion classification while suggesting avenues for further analysis, such as exploring biases and performance variations across different emotion classes.

Table 1. Results for class 0 – majority class

Model	Accuracy	Precision (%)	Recall (%)	F1-score (%)
LSTM	76	75	81	78
LR	60	59	86	70
DT	56	59	86	70
BERT	66	65	68	67

Table 2. Results for class 1 – minority class

Model	Accuracy	Precision (%)	Recall (%)	F1-score (%)
LSTM	76	74	66	70
LR	60	60	26	36
DT	56	60	26	36
BERT	66	66	63	65

Despite limitations, the study's findings offer insights into public sentiments and emotions related to cholera outbreaks as expressed on social media platforms.

4.2 Comparative Performance Analysis

Among the four models evaluated, LSTM had the highest accuracy and better precision, recall, and F1-score performance for both classes. Logistic Regression and Decision Tree models show similar results with the lowest accuracy and weaker performance in predicting the minority class (class 1). BERT offers a middle ground with moderate accuracy and relatively balanced precision, recall, and F1-scores. LSTM has the highest overall accuracy of 75%, and performs relatively well in terms of precision, recall, and F1-score for both classes.

The models have varying precision and recall values for different classes, which provide insights into their performance on each class. LSTM has balanced precision and recall values for both classes, indicating that it predicts both negative (class 0) and positive (class 1) instances well. Logistic Regression and Decision Tree have relatively high recall for class 0 but lower recall for class 1. This suggests they are better at identifying negative instances (class 0) than positive ones (class 1). The precision for class

1 is notably lower, implying that when they predict class 1, they tend to have a higher rate of false positives. BERT shows balanced precision and recall values for both classes, similar to the LSTM model. This indicates that it performs reasonably well in identifying both negative and positive instances.

The evaluation metrics for the Logistic Regression and Decision Tree models are identical, suggesting comparable performance between the two. Nonetheless, these models exhibit lower accuracy (0.60) in contrast to the LSTM model. On the other hand, the BERT model showcases a respectable accuracy score of 0.66. Notably, BERT also demonstrates balanced precision, recall, and F1-scores for both classes. Considering these factors, the chosen approach is to proceed with the LSTM model due to its higher accuracy and performance across various evaluation criteria. The emotion analysis conducted on the cholera X dataset through machine learning models has yielded findings regarding the emotional responses of individuals to the Hammanskraal cholera outbreak. The Long short-term memory (LSTM) model emerges as a standout performer, demonstrating a commendable ability to balance precision and recall for both negative and positive emotions. This suggests that LSTM effectively captures the nuances of emotion expressed in tweets related to the cholera situation, establishing it as a robust model for emotion classifications in this context. On the other hand, logistic regression and decision tree models face challenges, particularly in effectively predicting positive emotions. The lower F1-scores for class 1 (positive emotion) indicate that these models may struggle to capture the optimistic or supportive emotions related to the cholera outbreak. Further exploration and feature refinement might be necessary to enhance the performance of these models. In contrast, the Bidirectional Encoder Representations from Transformers (BERT) model exhibits a balanced performance with equally high F1-scores for both negative and positive emotions. This balanced performance underscores BERT's comprehensive understanding of the diverse emotions expressed in tweets related to the cholera outbreak. BERT's ability to capture the complexity and variability of emotion positions it as a suitable model for studying emotions in this domain.

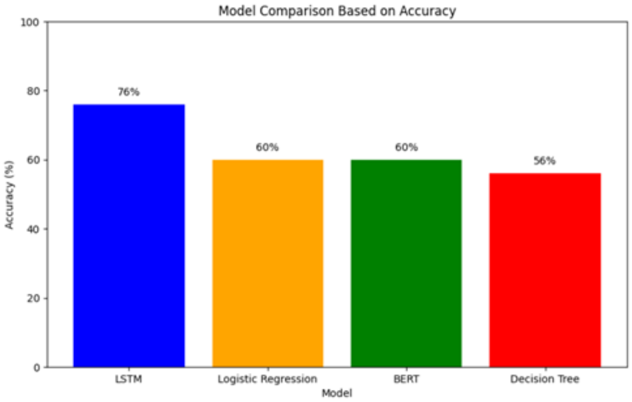


Fig. 5. Model performance based on accuracy

5 Conclusion

The emotion classification of social media text data connected to disease outbreaks contributes to a better understanding of people in such situations. Our study classified cholera tweets based on their emotional content. We extracted over 23000 tweets from X across several languages. The dataset was pre-processed before being labelled with NLTK's sentiment analyzer and separated into 80% training and 20% test data. Six (6) emotions as suggested by [44] were identified in our classification. Moreover, ML and DL models were developed for emotion categorization, using LSTM, LR, DT, and BERT. The models were trained on the cholera dataset were LSTM, LR, DT, and BERT produced a classification accuracy of 76%, 60%, 56%, and 66% respectively.

Using Ekman's basic emotions, the finding of the study reveals disgust as the dominant emotion expressed, accounting for a significant emotion observed in the dataset. Various factors can influence this emotion distribution, such as: (i) Public Perception: Cholera, which is one of the hashtags used to scrape tweets, is a serious and potentially life-threatening disease. Negative emotions may arise from fear, concern, or negative experiences related to Cholera and its impact on individuals, communities, or public health. (ii) Outbreak Context: The occurrence of a cholera outbreak can amplify negative emotions. During an outbreak, there may be heightened public attention, media coverage, and discussions focused on the negative aspects, such as the spread of the disease, its impact on affected areas, and the challenges in controlling and managing the outbreak. (iii) Emotional Impact: Diseases like Cholera can evoke strong emotional responses, particularly when they affect vulnerable populations or regions with limited healthcare infrastructure. Negative emotions may arise from empathy towards affected individuals, frustration with handling the outbreak, or anger towards perceived negligence or inadequate response measures. Considering these factors and interpreting the emotional results in the context of the specific dataset and analysis methodology used is important.

This provides a mechanism to gain insight into public perception and could be used for real-time monitoring of public emotion during events such as Cholera outbreaks. Future research could explore integrating different data modalities, such as images, videos, and user engagement metrics, with text-based emotion analysis.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

6 References

1. WHO: Cholera: key facts. (2023).
2. Cheng, X., Wang, Y., Huang, G.: Dynamics of cholera transmission model with imperfect vaccination and demographics on complex networks. J. Franklin Inst. 360, 1077–1105 (2023). https://doi.org/10.1016/j.jfranklin.2022.12.006.

3. Crooks, A.T., Hailegiorgis, A.B.: An agent-based modeling approach applied to the spread of cholera. Environ. Model. Softw. 62, 164–177 (2014). https://doi.org/10.1016/j.envsoft.2014.08.027.
4. National Institute for communicable Diseases: UPDATE | OUTBREAK OF CHOLERA IN SOUTH AFRICA (1 MARCH 2023). (2023).
5. Tulchinsky, T.H.: John Snow, Cholera, the Broad Street Pump; Waterborne Diseases Then and Now. In: Case Studies in Public Health. pp. 77–99. Elsevier (2018). https://doi.org/10.1016/B978-0-12-804571-8.00017-2.
6. Deen, J., Mengel, M.A., Clemens, J.D.: Epidemiology of cholera. Vaccine. 38, A31–A40 (2020). https://doi.org/10.1016/j.vaccine.2019.07.078.
7. Zerbo, A., Castro Delgado, R., González, P.A.: A review of the risk of cholera outbreaks and urbanization in sub-Saharan Africa. J. Biosaf. Biosecurity. 2, 71–76 (2020). https://doi.org/10.1016/j.jobb.2020.11.004.
8. Department of Health: Update on the response to Cholera Outbreak in South Africa. (2023).
9. UNICEF: Two children among latest cholera deaths in Gauteng province. (2023). https://doi.org/https://www.unicef.org/southafrica/press-releases/two-children-among-latest-cholera-deaths-gauteng-province.
10. Ngwa, M.C., Young, A., Liang, S., Blackburn, J., Mouhaman, A., Morris, J.G.: Cultural influences behind cholera transmission in the Far North Region, Republic of Cameroon: a field experience and implications for operational level planning of interventions. Pan Afr. Med. J. 28, 311 (2017). https://doi.org/10.11604/pamj.2017.28.311.13860.
11. Merten, S., Schaetti, C., Manianga, C., Lapika, B., Chaignat, C.-L., Hutubessy, R., Weiss, M.G.: Local perceptions of cholera and anticipated vaccine acceptance in Katanga province, Democratic Republic of Congo. BMC Public Health. 13, 60 (2013). https://doi.org/10.1186/1471-2458-13-60.
12. Liu, B.: Sentiment Analysis and Opinion Mining. Springer International Publishing, Cham (2012). https://doi.org/10.1007/978-3-031-02145-9.
13. Cambria, E., Schuller, B., Xia, Y., Havasi, C.: New Avenues in Opinion Mining and Sentiment Analysis. IEEE Intell. Syst. 28, 15–21 (2013). https://doi.org/10.1109/MIS.2013.30.
14. Ridzwan Yaakub, M., Iqbal Abu Latiffi, M., Safra Zaabar, L.: A Review on Sentiment Analysis Techniques and Applications. IOP Conf. Ser. Mater. Sci. Eng. 551, 012070 (2019). https://doi.org/10.1088/1757-899X/551/1/012070.
15. Wankhade, M., Rao, A.C.S., Kulkarni, C.: A survey on sentiment analysis methods, applications, and challenges. Artif. Intell. Rev. 55, 5731–5780 (2022). https://doi.org/10.1007/s10462-022-10144-1.
16. Bandopadhyaya, S.: Transcendence through social media. J. Media Commun. Stud. 8, 25–30 (2016). https://doi.org/10.5897/JMCS2015.0469.
17. Qualman, E.: Socialnomics: How social media transforms the way we live and do business. (2012).
18. Zhan, Y., Han, R., Tse, M., Ali, M.H., Hu, J.: A social media analytic framework for improving operations and service management: A study of the retail pharmacy industry. Technol. Forecast. Soc. Change. 163, 120504 (2021). https://doi.org/10.1016/j.techfore.2020.120504.

19. Wang, Z., Kim, H.G.: Can Social Media Marketing Improve Customer Relationship Capabilities and Firm Performance? Dynamic Capability Perspective. J. Interact. Mark. 39, 15–26 (2017). https://doi.org/10.1016/j.intmar.2017.02.004.
20. Kavanaugh, A., Fox, E.A., Sheetz, S., Yang, S., Li, L.T., Whalen, T., Shoemaker, D., Natsev, P., Xie, L.: Social media use by government. In: Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times. pp. 121–130. ACM, New York, NY, USA (2011). https://doi.org/10.1145/2037556.2037574.
21. Zaki, M.: Digital transformation: harnessing digital technologies for the next generation of services. J. Serv. Mark. 33, 429–435 (2019). https://doi.org/10.1108/JSM-01-2019-0034.
22. Gupta, S., Leszkiewicz, A., Kumar, V., Bijmolt, T., Potapov, D.: Digital Analytics: Modeling for Insights and New Methods. J. Interact. Mark. 51, 26–43 (2020). https://doi.org/10.1016/j.intmar.2020.04.003.
23. Du, R.Y., Netzer, O., Schweidel, D.A., Mitra, D.: Capturing Marketing Information to Fuel Growth. J. Mark. 85, 163–183 (2021). https://doi.org/10.1177/0022242920969198.
24. Burstein, P.: The Impact of Public Opinion on Public Policy: A Review and an Agenda. Polit. Res. Q. 56, 29–40 (2003). https://doi.org/10.1177/106591290305600103.
25. Bryson, J.M., Quick, K.S., Slotterback, C.S., Crosby, B.C.: Designing Public Participation Processes. Public Adm. Rev. 73, 23–34 (2013). https://doi.org/10.1111/j.1540-6210.2012.02678.x.
26. Jatla, S., Avula, D.: Hybrid Deep Model for Sentiment Analysis of COVID-19 Twitter Data. In: 2022 International Interdisciplinary Humanitarian Conference for Sustainability (IIHC). pp. 1616–1625. IEEE (2022). https://doi.org/10.1109/IIHC55949.2022.10060027.
27. Hossain, G.M.S., Asaduzzaman, S., Sarker, I.H.: A Deep Learning Approach for Public Sentiment Analysis in COVID-19 Pandemic. In: 2022 2nd International Conference on Intelligent Technologies (CONIT). pp. 1–6. IEEE (2022). https://doi.org/10.1109/CONIT55038.2022.9847839.
28. Singh, R., Singh, R., Bhatia, A.: Sentiment analysis using machine learning techniques to predict outbreaks and epidemics. (2018).
29. Kaur, H., Ahsaan, S.U., Alankar, B., Chang, V.: A Proposed Sentiment Analysis Deep Learning Algorithm for Analyzing COVID-19 Tweets. Inf. Syst. Front. 23, 1417–1429 (2021). https://doi.org/10.1007/s10796-021-10135-7.
30. Qin, Z., Ronchieri, E.: Exploring Pandemics Events on Twitter by Using Sentiment Analysis and Topic Modelling. Appl. Sci. 12, 11924 (2022). https://doi.org/10.3390/app122311924.
31. Thakur, N., Han, C.: An Exploratory Study of Tweets about the SARS-CoV-2 Omicron Variant: Insights from Sentiment Analysis, Language Interpretation, Source Tracking, Type Classification, and Embedded URL Detection. COVID. 2, 1026–1049 (2022). https://doi.org/10.3390/covid2080076.
32. Ogbuju, E., Oladipo, F., Yemi-Petters, V., Abdumalik, R., Olowolafe, T., Aliyu, A.: Sentiment Analysis of the Nigerian Nationwide Lockdown Due to COVID19 Outbreak. SSRN Electron. J. (2020). https://doi.org/10.2139/ssrn.3665975.
33. Ahmed, D., Salloum, S.A., Shaalan, K.: Sentiment Analysis of Arabic COVID-19

Tweets. In: Proceedings of International Conference on Emerging Technologies and Intelligent Systems. pp. 623–632 (2022). https://doi.org/10.1007/978-3-030-85990-9_50.

34. Duan, J., Zhang, P.-F., Qiu, R., Huang, Z.: Long short-term enhanced memory for sequential recommendation. World Wide Web. 26, 561–583 (2023). https://doi.org/10.1007/s11280-022-01056-9.

35. Okut, H.: Deep Learning for Subtyping and Prediction of Diseases: Long-Short Term Memory. In: Deep Learning Applications. IntechOpen (2021).

36. Mozafari, M., Farahbakhsh, R., Crespi, N.: A BERT-Based Transfer Learning Approach for Hate Speech Detection in Online Social Media. Presented at the (2020). https://doi.org/10.1007/978-3-030-36687-2_77.

37. Tahmid Rahman Laskar, M., Huang, J., Hoque, E.: Contextualized embeddings based transformer encoder for sentence similarity modeling in answer selection task. In: LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings. pp. 5505–5514 (2020).

38. Selva Birunda, S., Kanniga Devi, R.: A Review on Word Embedding Techniques for Text Classification. In: Innovative Data Communication Technologies and Applications. pp. 267–281 (2021). https://doi.org/10.1007/978-981-15-9651-3_23.

39. Feng, J., Xu, H., Mannor, S., Yan, S.: Robust logistic regression and classification. In: International Conference on Neural Information Processing Systems (2014).

40. Xu, X., Frank, E.: Logistic Regression and Boosting for Labeled Bags of Instances. In: Advances in Knowledge Discovery and Data Mining. pp. 272–281 (2004). https://doi.org/10.1007/978-3-540-24775-3_35.

41. Mor, A., Kumar, M.: Multivariate short-term traffic flow prediction based on real-time expressway toll plaza data using non-parametric techniques. Int. J. Veh. Inf. Commun. Syst. 7, 32 (2022). https://doi.org/10.1504/IJVICS.2022.120821.

42. Stiglic, G., Kocbek, P., Fijacko, N., Zitnik, M., Verbert, K., Cilar, L.: Interpretability of machine learning-based prediction models in healthcare. WIREs Data Min. Knowl. Discov. 10, (2020). https://doi.org/10.1002/widm.1379.

43. Sagi, O., Rokach, L.: Approximating XGBoost with an interpretable decision tree. Inf. Sci. (Ny). 572, 522–542 (2021). https://doi.org/10.1016/j.ins.2021.05.055.

44. Ekman, P., Friesen, W.: Constants across cultures in the face of emotion. J. Pers. Soc. Psychol. 17, 124–129 (1971).

Identifying Deep Learning Models for Detecting Child Online Threats to Inform Online Parents Education: A Systematic Literature

Jennyphar Kavikairiua[1][0009-0004-1142-5423], Fungai Bhunu Shava[2][0000-0002-6219-8206] and Mercy Chitauro[3][0009-0004-1478-5547]

[1, 2, 3] Namibia University of Science and Technology, Windhoek, Namibia
[1]kjennyphar@gmail.com
[2]fbshava@nust.na
[3]mchitauro@nust.na

Abstract. The COVID-19 pandemic has increased children's reliance on the Internet, exposing them to online risks, safety issues, mental health impacts, and educational inequities, and made it more difficult for parents to keep an eye on their children's online activities. Most parents are unaware of online risks, despite advice against strangers, sharing sensitive information, and monitoring social circles, but often overlook the importance of ensuring a safe online experience. This paper presents deep learning models that can be used to educate parents on online child protection. In this paper, a systematic literature review was used to explore deep learning models for detecting child online threats to inform online parents. Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) were identified as appropriate deep learning models in this study and will in the future be used as a reference point to develop a deep learning model that can educate parents on protecting their children online. This study will have a significant impact on the field of cyber security since it is envisaged that the new ways of analyzing threats and proposing solutions to equip parents to protect children online proposed in this paper will enable parents to monitor their children's exposure and take necessary precautions in relation to trending threats. The model can also enable tracking of children's online interactions, identifying patterns, detecting risks, and understanding changes in online threats like cyberbullying. It is also superior for speech synthesis, question answering, and language modelling, among others, thus, parents can ask questions.

Keywords: Deep Learning Model, Educate Parents, Online Child Protection.

1 Introduction

One in three children under 18 Internet users are unsupervised [1], while 79% of 15–24-year-olds use the Internet globally, a 14% increase from the general population [2]. Many machine learning techniques are being developed for tackling risk against children online, such as filtering content, controls by parents [1], flag harmful behaviour [3] and detection of child sexual exploitation. Such systems limit communication with children to keep them safe from child predators [1]. The COVID-19 pandemic has sped up the online transition in children, leading to increased dependency on digital platforms, distance learning issues, online risk exposure, safety concerns, mental health impacts, educational disparities, and difficulties in online child protection faced by parents.

According to a study done by [7], most parents are unaware of online risks and have no idea what their children are doing online. Additionally, literature shows a lot of parents had trouble juggling their work obligations, which prevented many of them from being able to devote time to their children at home [8]. Children are thus left alone to figure out how to adapt to life challenges including growing up online. Many parents counsel their children against answering the door while they are home alone, talking to strangers, or giving out sensitive information over the phone in the physical space. Parents also keep an eye on the places their children go, the friends they hang out with, and the media they read and watch. However, in the age we live in, digital tools have fundamentally altered both the world and children's experiences [9]. Young Internet users still have little cyber security awareness and understanding [10]. Just as it is in the physical world, children lack the necessary knowledge to completely comprehend the different online threats or how to avoid and resist them. For children to solicit for support they need to trust the support system. Children are less likely to speak up about potential or actual harm they are encountering when they lack a reliable support system and when there is more fear than trust of their own parents/ caregivers or guardians. Furthermore, many parents are not aware that managing and parenting a child and ensuring a safe online experience at the same level as in the physical space is necessary. Even though parents are currently not ensuring a safe online experience for children, the researchers propose that it is prudent to suggest a digital solution to solve a digital solution as the (old) adage says, “set a thief to capture a thief”. AI is used in education to enable students to customize their approaches to learning challenges based on their own distinct experiences and favorites [5]. And to adapt to each student's prior knowledge, rate of learning, and desired learning outcomes to maximize learning [5]. The researchers propose to explore suitable models for deep learning which could potentially be developed for educating parents on how to ensure a safe online experience for their children.

2 Objective

This study’s primary goal is to identify deep learning models to educate parents on online child protection. The following question was formulated to achieve the above-mentioned objective.

- What are the appropriate deep learning models to educate parents on online child protection?

3 Methodology

A systematic review of existing literature was conducted in this study to provide a wider overview of prospective research on deep learning techniques and online child protection. The following elements (inclusion criteria, exclusion criteria, and methods) were included for the reader to comprehend the systematic review's main purpose.

3.1 Inclusion Criteria

- All the articles selected are academic and blind peer- reviewed.
- The articles are related to online child protection and machine learning.
- The articles answered the research question.
- Any extra details that seem important and valuable were also chosen such as the types of deep learning networks.
- Articles written in English.

3.2 Exclusion Criteria

- Articles without a thorough discussion of online child protection and machine learning, regardless of whether they are scholarly or peer-reviewed, were not included.
- Articles that do not contain information on this study's indicated keywords were not considered.
- Any literature findings that contained a lot of repetition were discarded.
- Literature with insufficient settings or references was not considered.

3.3 Methods

The systematic literature review analysis focused on research articles from databases in computer science and information technology, specifically those addressing the application of machine learning in online child protection. The researchers conducted a comprehensive search across several peer-reviewed databases in the fields of Computer Science and Information Technology, including ProQuest, IEEE Xplore, ScienceDirect, SpringerLink, Scopus, and Google Scholar, aiming to address the specified research question. The search was executed on October 24, 2023, initially 298 publications were identified. A Microsoft Excel spreadsheet was used to systematically classify the relevant data from these publications, which included titles, abstracts, keywords, author names, journal names, and publication year. Following the initial screening, 222 articles were excluded. These excluded articles either lacked a comprehensive discussion of online child protection and machine learning or deep learning. Additionally, articles that failed to incorporate the specified keywords pertinent to this study were excluded. A full-text evaluations were performed on the remaining 76 academic papers, with strict eligibility criteria focusing only on topics related to online child protection and machine learning, which ultimately resulted in 19 relevant articles being included in the review. The databases search parameters were defined to include several pertinent keywords such as "deep learning" OR "deep neural network," AND "online child protection," Or "child online protection," OR "educating parents". Table 1. outlines the publisher-related literature relevant to the study.

Table 1. Publisher-related literature relevant to the study.

Databases	Relevant Literature		
	Downloaded	After analysing the abstract and title	After analysing the entire content
ProQuest	20	12	2
IEEE Xplore	112	44	11
ScienceDirect	63	5	1
SpringerLink	17	3	0
Scopus	51	12	5
Google Scholar	35	Duplicates 25 Removed and the other 10 were not relevant	0
TOTAL	**298**	**76**	**19**

Fig. 1 shows the total number of citations per year for publications cited in this paper related to machine learning in online child protection, from 2009 to 2023.

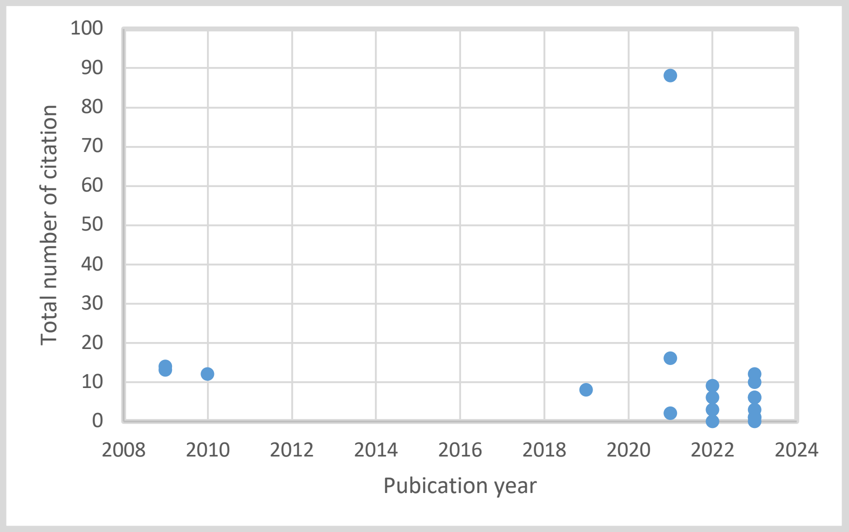


Fig. 1. Number of citations per year for publications

Fig. 1 illustrations that in the initial years (2009, 2010), publications were cited for an average of =10 citations each. There are no publications or citations between 2011 and

2018. The paper referenced the most in this review is that of 2021 with nearly 90 citations. The total citations for 2022 is 18 and the total for 2023 is 39. The recent rise in citations and publications from 2021- 2023 reflects technological advancements in machine learning and a growing need for online child protection solutions. A detailed summary of the findings from these papers is presented in the results section.

4 Results and Discussion

This section presents findings of relevant explanatory factors and explores their applicability to the deep learning model for educating parents on online child safety, as shown in Table 2.

Table 2. Online child protection, Techniques used and Applicability to DL Model.

Online child Protection	Data	Techniques used	Dataset Used	Applicability to DL Model	Source
Online harassment, Cyber-stalking, online grooming	Text	Decision trees, Naïve Bayes (NB), and Support Vector Machines (SVM)	Text messages	No	[12]
Online child grooming	Text	K-nearest neighbours classifier, SVM	Actual online child grooming and non-grooming conversations	No	[13]
Online grooming	Text	Bag of words (BoW), fuzzy-rough feature selection, and Fuzzy twin support vector machine.	Perverted justice (PJ) and PAN13	No	[14]
Cyber bullying and Online grooming	Text, Image	Adult image detecting algorithm, NLP, irrelevant post detection algorithm	Bad words datasets and sensitive word datasets	Yes	[15]
Detects port scan attempt	Port scans	Deep learning and SVM	CICIDS2017 dataset	Yes	[16]
Detect Cyber bullying on suspicious websites and blogs	Text	SVM	Data collected by Internet Patrol	No	[17]

Detect Cyberbully	Text	Instance-based learner and C4.5 decision tree learner	Formspring.me	No	[18]
Child sexual abuse material (CSAM)	Text	ML/DL, and NLP	Self-build, Third-party and Open datasets	Yes	[11]
Abusive language detection	Text	IBK, SVM, NB, JRip, Logistic, CLSTM, CNN, LSTM, and BLSTM	YouTube comments	Yes	[19]
Cyberbullying detection	Text	Logistic regression classification, Decision tree, NB algorithm, Random Forest, K-nearest neighbor, SVM,	Social Media Websites	No	[20]
Cyber security phishing detection	Text	SVM, Logistic regression, Boosted decision tree Averaged perceptron, Decision Forest. Neural network,	Phishing email collection, phishing legitimate full, spam or not spam dataset	Yes	[21]
Online grooming	Text	LDA model, and the life cycle of grooming	Real cyber paedophile chats downloaded from Perverted justice (PJ)	No	[22]
Detecting abusive messages in chat logs.	Text	NLP, user behavioural info, structure of conversations	Massively Multiplayer Online Role-Playing Game	No	[23]
Detection of Child Sexual Abuse (CSA), Pornography, and age	Images	Deep CNN	Pornographic-2M and Juvenile-80k	Yes	[24]
Online sexual predatory chats detection	Text	LIWC, term frequency-inverse document frequency (TFIDF), syntactical, sentiment polarity,	Chat-logs	No	[3]

		Markov chains, fuzzy-rough, SVM, BOW			
Fight child pornography in social media age	Images	CNN modelling	SEIC/non-SEIC content (Brazilian Federal Police)	Yes	[25]
Detecting Sexual Predatory Chats	Text	Artificial neural networks, Word2Vec	PAN 2012 dataset	Yes	[26]
Child Sexual Abuse Media (CSAM)	Image, video, PDF	CNN	Project VIC dataset	Yes	[27]
Online grooming detection	Text	SVM, K-NN, DT, Binary Logistic Regression, Naive Bayes, Maximum Entropy, EM and EMSIMPLE, Phrase matching, Rule-based techniques, Ring-based Classifier	PAN2012, MovieStarPlanet, Perverted Justice	No	[28]
Child safety and Protection in online gaming	Chat logs	Naïve Bayes (NB), DT, multilayer perceptron (MLP), k-nearest neighbour (k-NN) and SVM	Child paedophile chats were collected from PJ, and Sexual chats among adults were collected from Omegle data collection.	No	[1]
Facial Emotion recognition	Image	Multi Label Classification, CNN, DNN.	Kaggle dataset of 1857 autism spectrum disorder children and 1850 Typically Developed (TD) children	Yes	[29]
Emotion detection	Text based	NLP	ISEAR Dataset, SemEval, Emobank,	Yes	[30]

			EmoInt,Cecilia Ovesdotter Alms Affect data, Daily Dialog, AMAN’S Emotion dataset, Grounded Emotion Data, Emotion-Stimulus data, Crowdsourcing dataset, MELD dataset, Emotionlines and Smile dataset		
Child sentiment analysis	Text	NLP	Collect raw texts as data.	Yes	[31]
Detect online sexual predation.	Text	NLP	Perverted-justice, cybersex logs available online and the NPS chat corpus	Yes	[32]
Detect Child Sexual Exploitation Material (CSEM)	Text	Supervised machine learning approach	Dataset extracted from the file names and file paths	No	[33]
Video porn detect	Static and motion information	CNNs	ImageNet dataset, Pornography-800 and Pornography-2k dataset	Yes	[34]
Adult Content Detection	Videos	Local Receptive Field-Extreme Learning Machine (LRF-ELM) model	NPDI dataset	Yes	[35]
Text-based spam filters	Text	Deep Convolutional Neural Network (DCNN)	Image Spam Hunter Dataset (ISH), Improved Dataset and Dredze ImageSpam Dataset	Yes	[36]

Detecting different types of abusive contents	Text	LinearSVC, Logistic Regression (Logit), Multinomial Naïve Bayes (MNB), RF, ANN, RNN, and LSTM	Gathered data from public comment sections on various social sites and online resources	Yes	[37]
Detect personal information.	Text	Deep Generative adversarial networks	Synthetic dataset	Yes	[38]

All sources indicate that most machine learning and deep learning tools developed to protect children against online threats posed to them focus on detecting child sexual exploitation content through analyzing text. Eight of those tools were developed to detect children's exposure to online grooming and four of the reviewed articles developed tools to detect children's exposure to online pornography.

In general, the researchers found from the study that most existing studies in deep learning focused on detecting child sexual exploitation on social networking sites namely online grooming exemplified by [12]; [13]; [14]; [15]; [22]; [39]. Four of the reviewed articles [24]; [25]; [34]; [32] which amounts to 53% indicated child online pornography as another online threat posed to children that machine learning and deep learning tools are focused on detecting. Table 3 outlines the applicable models from the literature review outcome of the explanatory variables of interest, that only used deep learning techniques in online child protection.

Table 3. Online child protection and Deep learning techniques

Online Child Protection	**DL techniques/Model**	**Reference**
Online grooming/ Online sexual predators/ adult content/ sexual abuse	Natural Language Processing, local Receptive Field-Extreme Learning Machine; artificial neural networks,	[28]; [22]; [32]; [35]; [26]; [24]; [27]
Cyber bullying (Abusive language/ abusive messages/ content)	CNN, LSTM, BLSTM, CLSTM, RNN, Natural language processing,	[11]; [19]; [37]; [15]
Pornography	Deep CNN, CNN modelling, Natural language processing (NLP)	[34]; [24]; [25]
Personal information detection/	Deep Generative adversarial networks; neural networks	[38]
Port scan attempt detection	Support Vector Machine algorithm and deep learning	[16]
Text based spam filter/ phishing detection	Deep Convolutional Neural Network (DCNN)	[36]; [21]

Facial Emotion recognition/ detection\ sentiment analysis	Multi Label Classification, Convolutional Neural Networks, Deep Neural Networks, Natural language processing	[29]; [30]; [31]

From the analysis, it was established that most existing tools developed for online child protection use Natural Language Processing techniques and the most used deep learning model is the Convolutional Neural Networks algorithms followed by Long short-term memory to analyze and detect online threats posed to children. It was established that most deep learning models on online child protection focus on online grooming, cyberbullying, and pornography. In addition, the least online child protection detected threats were personal information detection, port scan attempt detection and text-based spam filter.

This study aimed to identify appropriate deep learning models that will aid in the development of a model that educates parents on protecting their children safely online. Given the growing number of children using the Internet and the possible risks posed to them, and the lack of parent's knowledge on cyber security, and the lack of existing tools that are tailored to educate parents on online child protection. It was determined that most deep learning models developed to protect children against online threats posed to them focus on detecting child sexual exploitation content through CNN and LSTM techniques. The CNN-LSTM is a convolutional neural network built on long short-term memory that performs sequence prediction tasks primarily with spatial input such as images, videos, or the temporal structure of words in a sentence, paragraph, or text [40]. In issues requiring the classification of temporal information, such as human activity detection, text classification, video classification, sentiment analysis, typhoon formation predicting, and arrhythmia diagnosis, the combination of CNNs and LSTM units has already produced promising results [41]. Therefore, this research concludes that the optimal deep learning models for COP parent education may be LSTM and CNN because:

- **CNN:** CNNs are adept at analyzing images and videos, making them useful for detecting inappropriate content, including child pornography materials, grooming material, and other inappropriate multimedia materials as they evolve. They can automatically detect and flag inappropriate content directly from the visuals they are exposed to. There is no need for pretraining but rather training on what it is exposed to, enabling parents to monitor their children's exposure and take necessary precautions informed of trending threats.
- **LSTM:** LSTM recurrent neural networks are effective in processing sequential data, enabling the tracking of children's online interactions, identifying patterns, detecting risks, and understanding changes in online threats like cyberbullying. It is also superior for speech synthesis, question answering, and language modelling, among others, thus, parents can ask questions.

Our research aims to develop a deep learning model to educate parents. This study has highlighted the effectiveness of CNNs and LSTMs in identifying online threats to children, providing valuable insights for enhancing parental awareness and protective measures in the digital world.

4.1 Proposed Deep Learning Model

The systematic literature review conducted for this study highlighted the effectiveness of CNNs and LSTMs in identifying online threats to children. As a result, the proposed deep learning model will incorporate these models, providing valuable insights for enhancing parental awareness and protective measures in the digital world. Fig. 2 outlines a case study demonstrating CNN and LSTM potent for educating parents on online child protection.

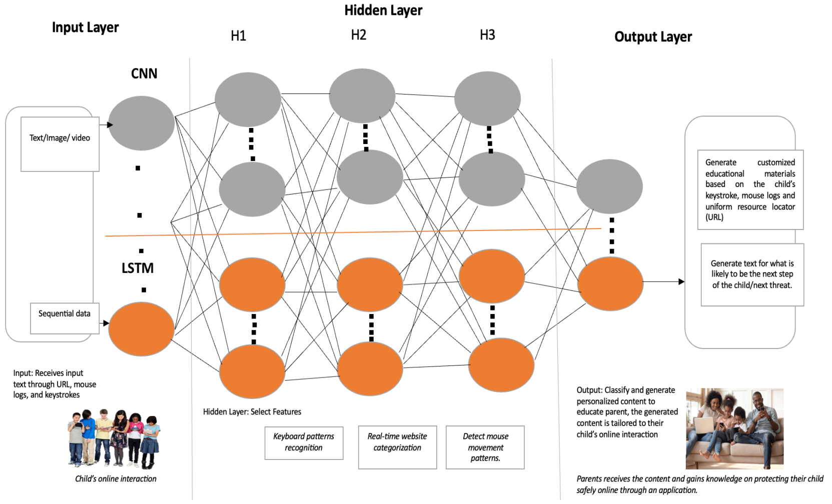

Hidden Layer
Input Layer
H1
H2
H3
Output Layer
CNN
Text/Image/ video
LSTM
Sequential data
Generate customized educational materials based on the child's keystroke, mouse logs and uniform resource locator (URL)
Generate text for what is likely to be the next step of the child/next threat.
Input: Receives input text through URL, mouse logs, and keystrokes
Child's online interaction
Hidden Layer: Select Features
Keyboard patterns recognition
Real-time website categorization
Detect mouse movement patterns.
Output: Classify and generate personalized content to educate parent, the generated content is tailored to their child's online interaction
Parents receives the content and gains knowledge on protecting their child safely online through an application.

Fig. 2. Proposed Deep Learning Model.

The deep learning model will receive input text through the Uniform Resource Locator (URL), mouse logs and keystrokes when a child interacts on a social networking site. It pays attention to things like the websites they visit, what they write, and even how they move the computer mouse and type on the keyboard. It uses two different methods to understand the child's online interaction. One method, called CNN, looks at the words, pictures, and videos the child sees online. It tries to find anything that might be bad or not suitable for them, like rude or harmful stuff. The other method, called LSTM, looks at how the child behaves online in a step-by-step way. It will figure out what the child might do next and if there's any danger in their online interaction. This information will be used to create special learning materials for the parent. Besides that, the model should also keep track on how the child is acting online to spot if something is not right, like if the child is being bullied or if they are facing problems. Then, it should give parents advice and information through a mobile application. The deep learning model will help parents understand their child's online activities and provides personalised guidance to protect them. As well as help parents monitor their child's online activities and protect them from potential online threats.

However, performance evaluation is crucial in deep learning processes to verify the effectiveness of the model. Reference [40] states that the accuracy and precision of multi-class classification models are calculated to evaluate their performance using confusion matrices, recall and f1 score. Accuracy is the ratio of correct predictions to total predictions across all classes, while precision is the ratio of true positives to the sum of true positives and false negatives [40], [42], [43]. Recall is the ratio of correctly predicted positive instances to all instances in a class [40], [42], [43]. Hence, the effectiveness of the proposed deep learning model will be evaluated using confusion matrices to determine accuracy, precision, recall, and F1-Score.

5 Conclusion

This study presents identified deep learning models that will aid in the development of a model that educates parents on protecting their children safely online, namely CNN and LSTM. An analysis on online child protection applicable deep learning techniques was carried out on literature from 2009 to 2023. This study identified CNN, LSTM, and NLP as the most used tools for online child protection, with LSTM and CNN being the best deep learning models for detecting child online threats mainly focusing on child sexual exploitation content. Parents can keep track of their children's exposure to inappropriate content by using CNNs to analyse photos and videos for cyber security. Children's Internet interactions can be monitored by LSTM recurrent neural networks, which can spot online trends and online dangers. The relevant deep learning models that have been identified in this study will be used as a guide the development a deep learning model that can detect child online threats to inform an application that can educate parents on the best ways to protect their children online. The PAN 2012, Bad words and the Perverted Justice datasets were identified as appropriate for training and testing the ensembled/ hybrid model. This work could significantly impact the field of cyber security since it will advance parents' understanding of online child protection and pave the way for further research into ways to protect children online.

References

[1] A. Faraz, J. Mounsef, A. Raza, and S. Willis, “Child Safety and Protection in the Online Gaming Ecosystem,” *IEEE Access*, vol. 10, pp. 115895–115913, 2022, doi: 10.1109/ACCESS.2022.3218415.

[2] International Telecommunication Union, “Measuring digital development Facts and Figures 2023,” ITU, 2023.

[3] C. H. Ngejane, J. H. P. Eloff, T. J. Sefara, and V. N. Marivate, “Digital forensics supported by machine learning for the detection of online sexual predatory chats,” *Forensic Science International: Digital Investigation*, vol. 36, no. March, 2021, doi: 10.1016/j.fsidi.2021.301109.

[4] World Economic Forum, “Artificial Intelligence for Children,” no. March, p. 38, 2022.

[5] H. Munir, B. Vogel, and A. Jacobsson, “Artificial Intelligence and Machine Learning Approaches in Digital Education: A Systematic Revision,” *Information (Switzerland)*, vol. 13, no. 4. MDPI, Apr. 01, 2022. doi: 10.3390/info13040203.

[6] E. Sánchez and M. Lama, “Artificial Intelligence and Education,” *Encyclopedia of Artificial Intelligence*, pp. 138–143, 2022, doi: 10.4018/978-1-59904-849-9.ch021.

[7] UNICEF, “The digital dance : parenting in an online world,” 2020, [Online]. Available: https://www.lifelinechildline.org.na/wp-content/uploads/2020/09/The-Digital-Dance-Parents-Booklet.pdf

[8] International Telecommunication Union (ITU), “Guidelines for parents and educators on Child Online Protection 2020,” ITU, 2020.

[9] T. Burns and F. Gottschalk, “What do we know about children and technology?,” *Educational Research and Innovation*, pp. 1–16, 2019.

[10] J. Kahimise and F. Bhunu Shava, “An analysis of children’s online activities and behaviours that expose them to cybercrimes,” in *27th Telecommunications Forum, TELFOR 2019*, Institute of Electrical and Electronics Engineers Inc., Nov. 2019. doi: 10.1109/TELFOR48224.2019.8971089.

[11] V. M. Ngo, C. N. Dang, H. Chi, C. Thorpe, and S. Mckeever, “Investigation, Detection and Prevention of Online Child Sexual Abuse Material: A Comprehensive Survey Science and Technology Application for Sustainable Development (STASD) Research Group,” 2023, doi: 10.20944/preprints202301.0046.v1.

[12] H. M. Al-Khateeb and G. Epiphaniou, “How technology can mitigate and counteract cyberstalking and online grooming,” *Computer Fraud and Security*, vol. 2016, no. 1, pp. 14–18, 2016, doi: 10.1016/S1361-3723(16)30008-2.

[13] F. E. Gunawan, L. Ashianti, S. Candra, and B. Soewito, “Detecting online child grooming conversation,” *Proceedings - 11th 2016 International Conference on Knowledge, Information and Creativity Support Systems, KICSS 2016*, 2017, doi: 10.1109/KICSS.2016.7951413.

[14] P. Anderson, Z. Zuo, L. Yang, and Y. Qu, “An Intelligent Online Grooming Detection System Using AI Technologies,” *IEEE International Conference on Fuzzy Systems*, vol. 2019-June, 2019, doi: 10.1109/FUZZ-IEEE.2019.8858973.

[15] A. Upadhyay, A. Chaudhari, Arunesh, S. Ghale, and S. S. Pawar, “Detection and prevention measures for cyberbullying and online grooming,” *Proceedings of the International Conference on Inventive Systems and Control, ICISC 2017*, pp. 1–4, 2017, doi: 10.1109/ICISC.2017.8068605.

[16] D. Aksu and M. A. Aydin, “Detecting Port Scan Attempts with Comparative Analysis of Deep Learning and Support Vector Machine Algorithms,” *International Congress on Big Data, Deep Learning and Fighting Cyber Terrorism, IBIGDELFT 2018 - Proceedings*, pp. 77–80, 2019, doi: 10.1109/IBIGDELFT.2018.8625370.

[17] M. Ptaszynski, P. Dybala, T. Matsuba, F. Masui, R. Rzepka, and K. Araki, “Machine learning and affect analysis against cyber-bullying,” *Proceedings of the 1st International Symposium*

on Linguistic and Cognitive Approaches to Dialog Agents - A Symposium at the AISB 2010 Convention, no. May 2014, pp. 7–16, 2010.

[18] K. Reynolds, A. Kontostathis, and L. Edwards, "Using Machine Learning to Detect Cyberbullying," 2021. [Online]. Available: www.noswearing.com.

[19] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. AbdelMajeed, and T. Zia, "Abusive language detection from social media comments using conventional machine learning and deep learning approaches," *Multimed Syst*, vol. 28, no. 6, pp. 1925–1940, 2022, doi: 10.1007/s00530-021-00784-8.

[20] M. A. Al-Garadi *et al.*, "Predicting Cyberbullying on Social Media in the Big Data Era Using Machine Learning Algorithms: Review of Literature and Open Challenges," *IEEE Access*, vol. 7, no. May, pp. 70701–70718, 2019, doi: 10.1109/ACCESS.2019.2918354.

[21] A. Mughaid, S. AlZu'bi, A. Hnaif, S. Taamneh, A. Alnajjar, and E. A. Elsoud, "An intelligent cyber security phishing detection system using deep learning techniques," *Cluster Comput*, vol. 25, no. 6, pp. 3819–3828, 2022, doi: 10.1007/s10586-022-03604-4.

[22] P. Zambrano *et al.*, "Technical mapping of the grooming anatomy using machine learning paradigms: An information security approach," *IEEE Access*, vol. 7, pp. 142129–142146, 2019, doi: 10.1109/ACCESS.2019.2942805.

[23] N. Cecillon *et al.*, "Graph embeddings for Abusive Language Detection To cite this version : HAL Id : hal-03042171," 2021.

[24] A. Gangwar, V. González-Castro, E. Alegre, and E. Fidalgo, "AttM-CNN: Attention and metric learning based CNN for pornography, age and Child Sexual Abuse (CSA) Detection in images," *Neurocomputing*, vol. 445, no. March, pp. 81–104, 2021, doi: 10.1016/j.neucom.2021.02.056.

[25] P. Vitorino, S. Avila, M. Perez, and A. Rocha, "Leveraging deep neural networks to fight child pornography in the age of social media," *J Vis Commun Image Represent*, vol. 50, no. December, pp. 303–313, 2018, doi: 10.1016/j.jvcir.2017.12.005.

[26] P. R. Borj, K. Raja, and P. Bours, "Detecting Sexual Predatory Chats by Perturbed Data and Balanced Ensembles," *Lecture Notes in Informatics (LNI), Proceedings - Series of the Gesellschaft fur Informatik (GI)*, vol. P-315, pp. 245–252, 2021.

[27] M. Pereira, R. Dodhia, H. Anderson, and R. Brown, "Metadata-Based Detection of Child Sexual Abuse Material," pp. 1–11, 2020, [Online]. Available: http://arxiv.org/abs/2010.02387

[28] P. R. Borj, K. Raja, and P. Bours, "Online grooming detection: A comprehensive survey of child exploitation in chat logs," *Knowl Based Syst*, vol. 259, p. 110039, 2023, doi: 10.1016/j.knosys.2022.110039.

[29] T. L. Praveena and N. V. M. Lakshmi, "Multi Label Classification for Emotion Analysis of Autism Spectrum Disorder Children using Deep Neural Networks," *Proceedings of the 3rd International Conference on Inventive Research in Computing Applications, ICIRCA 2021*, pp. 1018–1022, 2021, doi: 10.1109/ICIRCA51532.2021.9545073.

[30] F. A. Acheampong, C. Wenyu, and H. Nunoo-Mensah, "Text-based emotion detection: Advances, challenges, and opportunities," *Engineering Reports*, vol. 2, no. 7, pp. 1–24, 2020, doi: 10.1002/eng2.12189.

[31] E. A. E. Lucky, M. M. H. Sany, M. Keya, S. A. Khushbu, and S. R. H. Noori, "an Attention on Sentiment Analysis of Child Abusive Public Comments Towards Bangla Text and Ml,"

2021 12th International Conference on Computing Communication and Networking Technologies, ICCCNT 2021, 2021, doi: 10.1109/ICCCNT51525.2021.9580154.

[32] D. Bogdanova, P. Rosso, and T. Solorio, "On the Impact of Sentiment and Emotion Based Features in Detecting Online Sexual Predators," *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, no. July, pp. 110–118, 2012.

[33] W. Al Nabki, E. Fidalgo, and E. Alegre, "Short Text Classification Approach to Identify Child Sexual Exploitation Material Short Text Classification Approach to Identify Child Sexual Exploitation Material," no. October, 2020.

[34] M. Perez *et al.*, "Video pornography detection through deep learning techniques and motion information," *Neurocomputing*, vol. 230, no. July 2016, pp. 279–293, 2017, doi: 10.1016/j.neucom.2016.12.017.

[35] N. AlDahoul, H. A. Karim, M. H. L. Abdullah, M. F. A. Fauzi, S. Mansour, and J. See, "Local Receptive Field-Extreme Learning Machine based Adult Content Detection," IEEE, 2019.

[36] S. Srinivasan *et al.*, "Deep Convolutional Neural Network Based Image Spam Classification," in *Proceedings - 2020 6th Conference on Data Science and Machine Learning Applications, CDMA 2020*, Institute of Electrical and Electronics Engineers Inc., Mar. 2020, pp. 112–117. doi: 10.1109/CDMA47397.2020.00025.

[37] E. A. Emon, S. Rahman, J. Banarjee, A. K. Das, and T. Mittra, "A Deep Learning Approach to Detect Abusive Bengali Text," 2019.

[38] R. M. Alguliyev, F. J. Abdullayeva, and S. S. Ojagverdiyeva, "Protecting children on the internet using deep generative adversarial networks," *International Journal of Computational Systems Engineering*, vol. 6, no. 2, p. 84, 2020, doi: 10.1504/ijcsyse.2020.111207.

[39] P. R. Borj, K. Raja, and P. Bours, "Online grooming detection: A comprehensive survey of child exploitation in chat logs," *Knowl Based Syst*, vol. 259, p. 110039, 2022, doi: 10.1016/j.knosys.2022.110039.

[40] K. Yousaf and T. Nawaz, "A Deep Learning-Based Approach for Inappropriate Content Detection and Classification of YouTube Videos," *IEEE Access*, vol. 10, pp. 16283–16298, 2022, doi: 10.1109/ACCESS.2022.3147519.

[41] F. Madaeni, K. Chokmani, R. Lhissou, S. Homayouni, Y. Gauthier, and S. Tolszczuk-Leclerc, "Convolutional neural network and long short-term memory models for ice-jam predictions," *Cryosphere*, vol. 16, no. 4, pp. 1447–1468, Apr. 2022, doi: 10.5194/tc-16-1447-2022.

[42] I. Priyadarshini and C. Cotton, "A novel LSTM–CNN–grid search-based deep neural network for sentiment analysis," *Journal of Supercomputing*, vol. 77, no. 12, pp. 13911–13932, Dec. 2021, doi: 10.1007/s11227-021-03838-w.

[43] A. Roshanzamir, H. Aghajan, and M. Soleymani Baghshah, "Transformer-based deep neural network language models for Alzheimer's disease risk assessment from targeted speech," *BMC Med Inform Decis Mak*, vol. 21, no. 1, Dec. 2021, doi: 10.1186/s12911-021-01456-3.

The integration of Software Engineering into IT Degree Programs: An analysis of the curriculum and industry relevance

Alfred Hove Mazorodze https://orcid.org/0000-0002-4147-454X
Faculty of Information Technology
Belgium Campus ITversity
Pretoria, South Africa

E-mail: Mazorodze.a@belgiumcampus.ac.za

Abstract

Software Engineering is a core module in Information Technology (IT) degree programs. Because IT is ever evolving, the curriculum for Software Engineering should be updated more regularly to meet these technological changes. This research evaluates the Software Engineering curriculum at a selected private higher education institution in South Africa to fully understand the relevance of the content taught from a Software Engineer perspective. The study conducted a comprehensive review of the Software Engineering content taught at a selected private higher education institution in South Africa. Moreover, the study analyses the pedagogical approaches used to teach and assess Software Engineering students. Most importantly, the study solicits feedback through an online survey from the software engineering professionals in industry to ensure that graduates are well-equipped with the knowledge required to tackle problems in the real-world environment. A total of thirty-six (36) Software Engineers in South Africa took part in this study. Some of the current trends in Software Engineering include DevOps, Edge Computing, Microservices Architecture, Serverless Computing, Containerisation and Orchestration. The study established that technologies mentioned here should form part of the curriculum. The results of this study therefore offer insights for curriculum designers, educators, and policymakers to optimise the effectiveness of Software Engineering as a core module in Information Technology degrees. The study highly recommends that the curriculum be updated to incorporate DevOps, the Microservices Architecture as well as Serverless Computing among other contemporary technologies.

Keywords:
Software Engineering, Information Technology, curriculum review, pedagogical approaches, DevOps, Serverless Computing, Microservices Architecture, higher education.

1. Introduction

Software Engineering (SE) is a process of analysing user requirements and then designing, building, testing, deploying, and maintaining a software application which satisfies the user requirements[1]. The Software Engineering process follows a systematic approach from development up to maintenance of the software product. Software Engineering is taught as a module building towards an Information Technology (IT) degree at most tertiary institutions in South Africa and beyond. The Software Engineering curriculum should therefore be tailored and reviewed more regularly to match the dynamic world of Information Technology[2]. The goal of curriculum design is to define the learning outcomes that students are expected to achieve[3]. Curriculum includes content, instructional strategies, assessment methods as well as resources used for teaching and learning. The IEEE Software Engineering Competency Model (SWECOM) outlines competencies expected from a Software Engineering professional.

Since Software Engineering is an amalgamation of multiple modules inclusive of computer programming, systems analysis, systems design, and software testing[4], several tools are used for successful delivery of the module. Based on the specific requirements of the project, different tools and technologies are used to develop robust software. Some of the most common tools for Software Engineering include programming languages, version control systems, testing tools, project management and automation tools among others[5]. All tools and technologies mentioned here are taught as topics in the Software Engineering course and these are not the end-all. It is important to note that other modern technologies are introduced from time to time and the curriculum should be reviewed regularly to meet these important changes as confirmed by[2].

According to[6], the objectives of Software Engineering seek to make sure that the software developed is maintainable, efficient, correct, reusable, testable, reliable and portable. From the submission by[6], it becomes clear that the software developed should continuously evolve to meet the changing requirements. Efficiency is also an important aspect which ensures that the software should not waste computing resources such as memory and processing power, as these have a direct influence on the computer's performance and the overall ability to execute tasks. Reliability is an attribute of software quality which describes the extent to which a program is expected to perform its desired functions over a period. Due to the continued increase in the number of platforms and devices, it is vital to develop portable software applications. Portability is an important aspect of Software Engineering where the software developed can be transferred from one computer system to another. Therefore, a Software Engineering student should be able to develop portable and maintainable systems and master all these skills before entering the industry.

The need to review a Software Engineering curriculum is dependent on the emergence of new technologies and shifts in industry trends[2]. The Software Engineering curriculum guidelines by[7] serve as the basis to curriculum design. Other bodies responsible for Software Engineering curriculum are: Association of Computer Machinery (ACM) and the Accreditation Body for Engineering and Technology (ABET). It is advisable to regularly assess the curriculum to ensure its continued relevance, especially in IT. Numerous higher education institutions strive to update their Software Engineering curriculum every 3 to 5 years because of technology evolution and changes in industry trends. Because there should be an alignment of the curriculum and the IT industry expectations, the study therefore sought to answer the following questions:

Research questions:

- How is the Software Engineering curriculum structured at a selected private higher education institution?
- Which pedagogical approaches and assessment tools are used to assess Software Engineering students?
- What are the industry expectations of a Software Engineering graduate student?

2. Literature review

The literature review assesses the topics taught in Software Engineering at a private higher education institution and points to some modern topics not included in the current curriculum. The missing areas are then investigated in a survey to understand their relevance in the Software Engineering domain. The Software Engineering process has multiple stakeholders. Figure 1 below shows the Systems Development Life Cycle (SDLC), which is used as a fundamental tool to understand software development. The phases highlighted here call for different stakeholders, different roles, different skills, and different expectations. The SDLC form an integral part of the Software Engineering process.

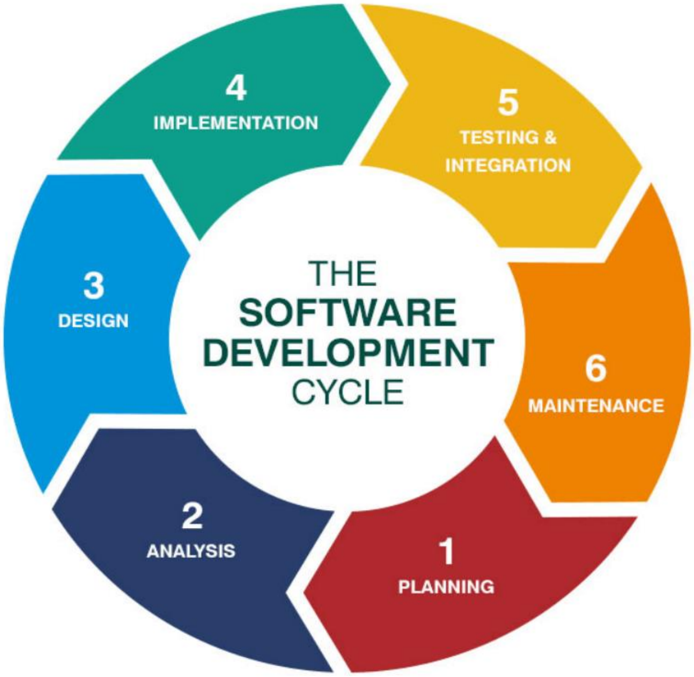


Figure 1: Systems Development Life Cycle (Adapted from[8])

The phases presented above are often iterative, meaning that the process may cycle back to earlier phases as and when necessary. In the curriculum, a Software Engineering student should understand all these phases in detail to excel at work. Like any other project, the planning phase is crucial to ensure the success of the software project. During this important phase, the objectives of the project are identified[8], the scope of the project is defined, and the resources required for the specific project are estimated[6]. This is an important phase for every aspiring Software Engineer. Many scholars[9],[10] concur that the analysis phase identifies and documents the needs and expectations of all stakeholders and end-users of a software system. This is a critical phase of Software Engineering, as it sets the foundation for the design and development of the software product.

During the design phase, the software design is broken down into smaller components[10], and the architecture of the software is determined. The design phase involves creating detailed specifications for the software, including data structures, algorithms, and user interfaces. The design phase also involves choosing the software development tools, technologies, and programming languages to be utilised. During the implementation phase, the software is coded, tested, and integrated into a complete system. Different programming languages such as Java, C++ and C# are used in this phase. Testing and integration are two critical processes in Software Engineering that ensure the quality and functionality of software products. Testing is the process of evaluating a software product or system to determine whether it meets the specified requirements and works as expected[9]. Maintenance is an essential part of Software Engineering that involves making changes and updates to software products after they have been released[10], and this is an ongoing process. Most of the Software Engineer's job involves system maintenance. Therefore, the Software Engineering curriculum should be updated to involve more of software maintenance tasks.

The Software Engineering curriculum also focuses on low-level design where the system architecture is broken down into smaller pieces that can be implemented by the software developers. Low-level design describes every module in detail by incorporating the logic behind every component in the system[11]. The design specifications are typically created by software architects or senior developers, in collaboration with the development team. Low-level design helps to ensure that the software is designed to meet the functional and non-functional requirements. An application created with poor design is difficult to maintain[12]. Meaning a good object-oriented design should be reusable, extensible, and maintainable and design patterns are helpful in trying to achieve all this. Design patterns provide software developers with a toolkit for handling problems that have already been solved[13]. Some of the key aspects of programming include coding, testing, documentation, collaboration and version control. Software Engineers should have a thorough understanding of various types of programming languages to cater for different software development needs. At this juncture, a Software Engineering graduate should be familiar with some of the languages presented below:

Object-Oriented Languages • C# • Java	**Software Engineer**	Scripting Languages • Python • JavaScript
Web Development Languages • HTML • CSS		Data Manipulation Languages • SQL • R

Figure 2: Programming Languages (Author's own creation)

Figure 2 shows some of the programming competencies expected from a Software Engineering graduate. It is important to note that this list is not exhaustive, and there are many more programming languages out there. Each language has its own strengths, weaknesses, and specific areas of application. It is vital to underscore that Software Engineers should have knowledge of many programming languages and should be willing to learn new languages all the time. The Software Engineering curriculum also entails an understanding of the layered architecture which seek to organize the components of a system into distinct layers based on their responsibilities and functionality. Each layer in the architecture provides a specific set of services and interacts with adjacent layers in a predefined manner. The main goal of layered architecture is to create a separation of concerns and promote modular design[6], allowing for easier development, maintenance, and scalability of the software system. The layered architecture is typically divided into three layers: presentation layer, business logic layer and the data access layer.

User Experience (UX) is the process of designing and improving the overall experience that users have while interacting with a software application. UX design is a very important aspect of the Software Engineering curriculum focusing on understanding users' needs, and preferences. UX encompasses various aspects in software development including usability, accessibility, visual design, and user engagement[14]. By incorporating UX principles and practices into Software Engineering, developers can create software applications that are not only functional and technically sound but also deliver exceptional user experiences. A well-designed user experience can lead to increased user satisfaction, improved adoption rates, and an overall success for the software product.

Software Testing is taught as an important topic in the Software Engineering module. The software testing process identifies the correctness, completeness, and quality of the software product[5]. The process involves a set of activities conducted with the intent of finding errors in software so that they can be corrected before the product is released to end users. A software product should only be released after it has gone through a proper process of development, testing and bug fixing[15]. Software testing looks at performance, stability and error handling conducted by software testers who are the professionals who perform testing by setting up test scenarios under controlled conditions and assessing the results. Software is considered of good quality if it meets the user requirements. Software Engineers should also have a thorough knowledge of the software testing process.

Literature review and content review has revealed that most of the contemporary topics in Software Engineering do not form part of the curriculum. Some of the current trends in Software Engineering include DevOps[16], Edge Computing[17], Microservices[18], Serverless Computing[19], Containerisation and Orchestration[20]**.** DevOps is a process that focuses on automating software development processes for increased efficiency and reliability. DevOps is one of the most modern software development methodologies that includes a set of practices and tools to integrate and automate software development. As alluded by[21], Edge Computing technologies seek to develop software solutions that leverage computing capabilities for faster processing and reduced latency. The Microservices Architecture is also another recent development in the Software Engineering domain seeking to implement software systems as a collection of small, independent services. Microservices provide quicker deployment time and can easily scale. Scalable software lowers the possibility of downtime[18] ensuring that the system can handle increased loads without performance degradation..

Serverless Computing develops and deploys applications without managing the server infrastructure directly[19]. For efficient deployment, scaling and management of applications, modern technologies like containers can be utilised. Docker and Kubernetes are some of the common container orchestration technologies which Software Engineering students must be taught. It is therefore important to reassess the current curriculum and proffer some recommendations for improvement, based on empirical data from Software Engineers in the field. The methodology adopted to complete this study is presented next.

3. Research Methodology

To assess the curriculum and pedagogical approaches to Software Engineering, an examination of the syllabus and course content was done. After the examination of institutional documents, an online survey was conducted with Software Engineers in South Africa to give their input on the relevance of different trends and technologies in the Software Engineering domain. The inputs from these Software Engineers are reported in this article and could help curriculum developers during the curriculum review process. The online questionnaire was pre-tested before administration to ensure that the questions were well-understood before distribution to the participants, a practice supported by many researchers[22].

The quantitative data from the Software Engineers was analysed collectively using Microsoft Office Excel, a researcher's statistical package of choice. The researcher maintained the anonymity of all research participants as stressed by[23] on the ethical principles in research. All the participants were informed about the aims and objectives of the study, and they participated out of their own volition. The researcher was granted permission to conduct this study in 2024 with reference BCI/2024/001.

4. Findings and discussion

The findings in this study are structured according to the study's three objectives.

4.1 Structure of the Software Engineering curriculum

The current Software Engineering curriculum at the selected private higher education institution is still relevant. The curriculum is structured into theoretical and practical concepts. Most of the content covered in the curriculum conform to the Software Engineering Competency Model[24] which seek to set standards and competencies of Software Engineers when developing and modifying software. Within the curriculum, some of the theoretical aspects taught include requirements analysis, software design, programming languages and agile methodologies among others. The practical concepts include the design of interfaces, coding, database development and deployment of applications. According to[6], every Software Engineer should know how to code and deploy applications using different technologies. However, there are more modern tools and technologies like Serverless Computing and DevOps in the Software Engineering domain which are not part of the current curriculum at the selected private higher education institution.

4.2 Pedagogical approaches and assessment tools for Software Engineering students

Teaching Software Engineering requires a combination of various pedagogical approaches to impart both knowledge and skills[24]. Some of the pedagogical approaches deployed involve project-based learning, lecture-based learning and flipped classrooms. Considering that Software Engineering is a practical module, project-based learning presents students with real-world problems that they must solve. In this form of pedagogy, students engage in the learning process by solving practical problems. The traditional lecture-based teaching remains effective for successful delivery of the Software Engineering principles. To keep students engaged, modern interactive elements and multimedia are recommended. Student engagement enhances learning and promotes the development of critical skills required by Software Engineers[25]. [26] argues that learning should be interactive, flipped classrooms are utilised to allow more interactive sessions. In a flipped classroom, students review course content before the class in a collaborative manner and hands-on exercises.

Evidence from the institutional documents established class tests, examinations, practical projects and presentations as the assessment tools used to assess Software Engineering students. In general, class tests and examinations are designed to evaluate understanding of core concepts and principles. These assessments could be in the form of multiple-choice questions, short questions and most importantly practical questions. Continuous assessments and oral examinations also form part of the assessment process. As alluded by[6], practical projects provide hands-on activities to solve problems using a range of tools and technologies. As part of the assessment process, every Software Engineering student should present his/her final project to an academic audience. This presentation is very important because in a real-life situation, after developing a solution to a problem, the developer should present this to the management and to the client for approval. It is important that all assessments align with the learning objectives and focus on both theoretical and practical competencies.

4.3 Industry expectations from Software Engineering graduates

The last objective of this study sought to understand the industry expectations from Software Engineering graduates. Based on the empirical evidence from the Software Engineers in South Africa, the industry expects graduates to possess a number of practical skills, inclusive of version control systems, agile methodologies, DevOps, web development and cloud computing skills. In South Africa, graduates from private institutions can be employed by either the private or the public sector. The following breakdown was obtained in relation to the companies.

4.3.1 Distribution of Software Engineers by company type in South Africa

A total of 36 Software Engineers took part in this investigation. Interestingly, 90% of the participants work in the private sector while the other 10% work for public enterprises. The breakdown is shown on Figure 3 below.

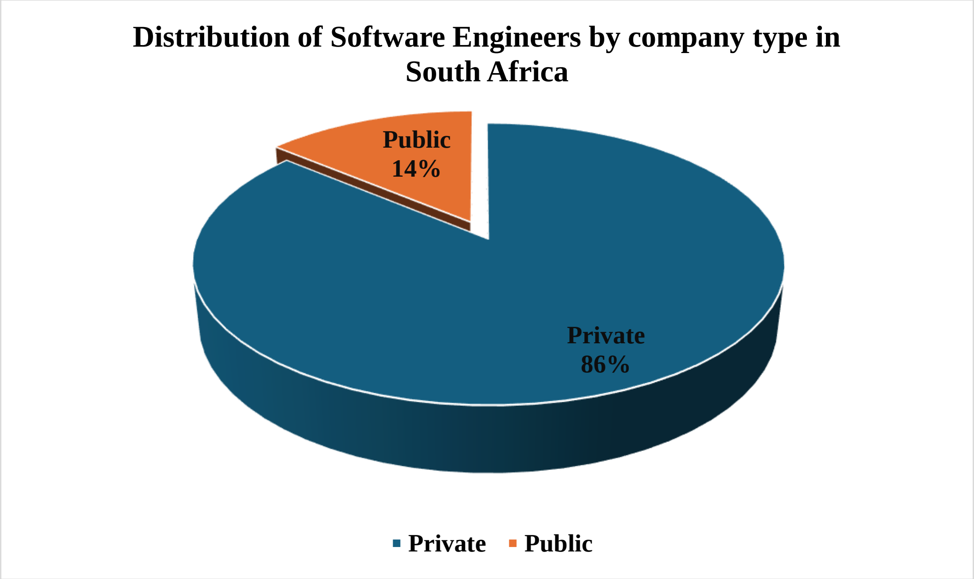


Figure 3: Distribution of Software Engineers by company type

Based on the empirical evidence presented above, we can therefore settle on the conviction that most Software Engineers in South Africa work for private companies. These Software Engineers had different years of work experience as explained in the subsequent section.

4.3.2 Work Experience

A total of thirty-six (36) responses were obtained with varying levels of experience as shown in Table 1 below:

Table 1: Work experience of the Software Engineers

Variable	Variable category	Frequency	Percentage
Working experience	Less than 1 year	1	3%
	1 year – 3 years	4	12%
	4 years – 10 years	19	53%
	More than 10 years	12	32%

Most of the participants had relevant work experience of between 4 and 10 years, with a 53% representation. On the same variable of work experience, exactly 32% of the participants had more than 10 years of work experience. It also emerged that 12% of the participants had between 1 and 3 years of related work experience. Only 3% of the participants had less than a year in the field. In this case, a relevant work experience is important to provide appropriate industry insights which could be used to improve the current curriculum at the specific higher education institution.

4.3.3 Software Engineering skills

The following Software Engineering skills are expected from IT graduates, as highlighted by the Software Engineers in practice. The information is presented descriptively using a bar graph.

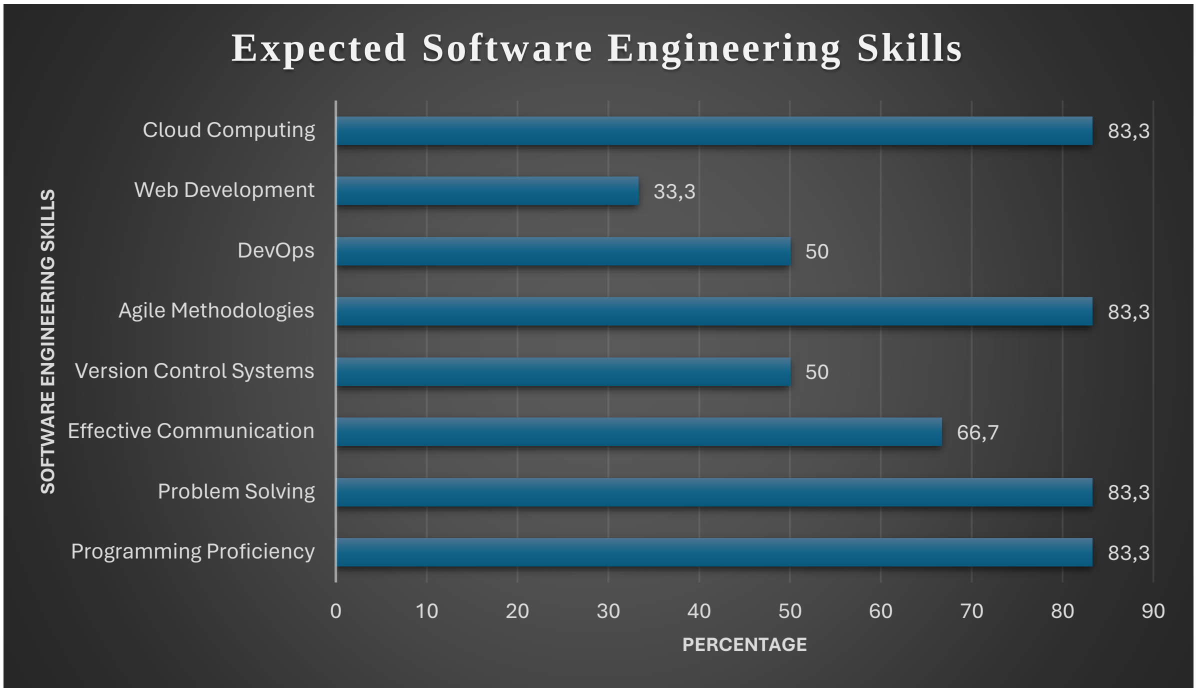


Figure 4: Software Engineering skills

As shown in figure 4, Cloud Computing, Agile Methodologies, problem solving and programming proficiency are the top skills expected from Software Engineering graduates. In addition to that, effective communication is key in the Software Engineering domain with a 66.7% representation. Other important skills include version control systems, DevOps, and web development. Therefore, we can extrapolate that Software Engineering demands a diverse set of skills. The survey sought to understand the relevance of certain technological developments in the Software Engineering domain. A Likert scale questionnaire elicited the following responses from 36 Software Engineers in South Africa.

4.3.4 Relevance of the technological trends in Software Engineering

Most technological trends in Software Engineering centre around automating processes, scaling operations, enhancing agility, and prioritizing user needs. These trends are spurred by advancements in technology and changing business demands. To stay competitive, Software Engineers must be updated on these modern technologies. Table 2 examines some of the emerging trends.

Table 2: Relevance of the technological trends in Software Engineering **(n = 36)**

Technological Trend	**Strongly Agree 5**	**Agree 4**	**Not Sure 3**	**Disagree 2**	**Strongly Disagree 1**
	Positive		**Neutral**	**Negative**	
DevOps and Continuous Integration	58.3%		13.9%	27.8%	
Microservices Architecture	66.7%		19.4%	13.9%	
Containerisation and Orchestration	88.9%		11.1%	0.0%	
Machine Learning (ML) and Artificial Intelligence (AI)	88.9%		5.6%	5.6%	
Serverless Computing	72.2%		27.8%	0.0%	
Edge computing	63.9%		13.9%	22.2%	

The empirical findings presented in Table 2 above is analysed using a bar graph as shown below. In Figure 5, the caption "positive" denotes participants who expressed strong agreement or agreement with the relevance of a specific technology. "Neutral" implies participants who were neither aligned nor decisive regarding a specific technology proposed. "Negative" characterizes those participants who disagreed or strongly disagreed with the significance of a technology in relation to their work. Since the study wants to establish the relevance of these technologies in the Software Engineering domain, the focus is therefore on the positive responses only. The study therefore analyses the relevance of the technologies using a bottom-up approach.

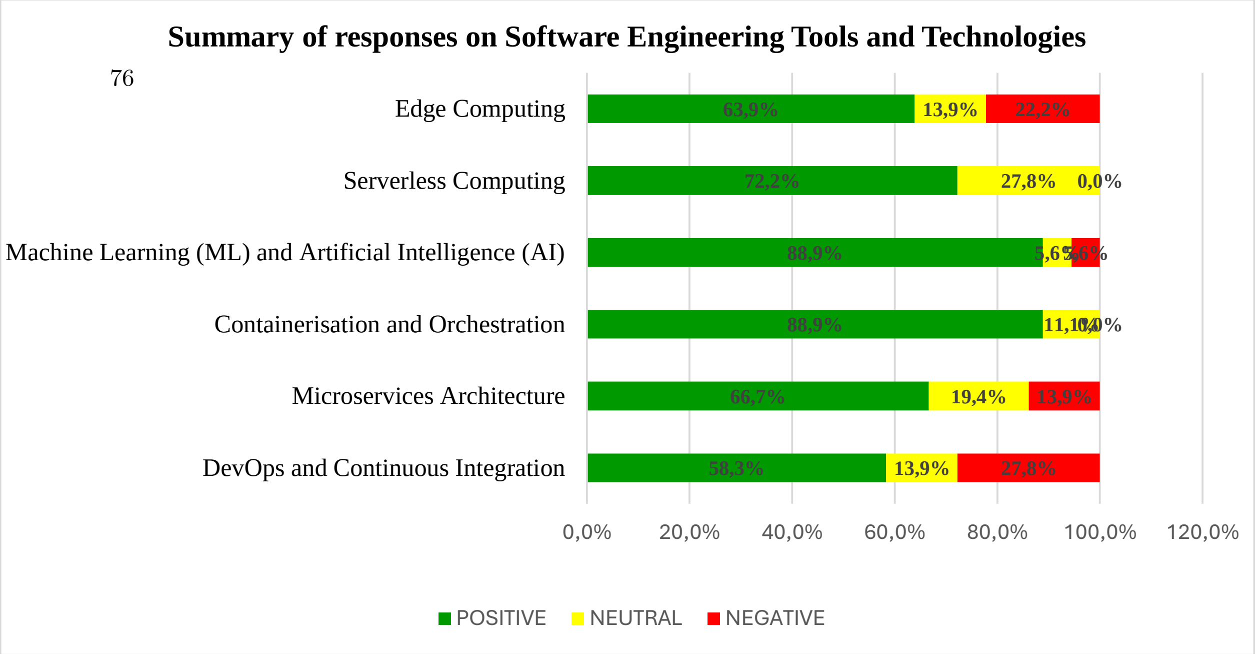


Figure 5: Technological trends in Software Engineering

DevOps and continuous integration were ranked as important trends used in Software Engineering. As evident in Figure 5 above, 58.3% of the participants submitted that DevOps and continuous integration is very relevant in the Software Engineering domain. On the same premise, 13.9% were non-aligned on DevOps and continuous integration. It was also submitted by 27.8% of the participants that DevOps and continuous integration is not relevant in the execution of their duties. DevOps and continuous integration focus on automating the software development processes for increased efficiency and reliability. Literature establishes that DevOps provides faster and better product delivery[27]. Therefore, it is important that students in higher education institutions are taught these practices, as they offer greater automation in more stable operating environments. This study therefore recommends the introduction of DevOps into the Software Engineering curriculum.

Along the same spectrum of technologies used for Software Engineering, the Microservices Architecture is also gaining popularity in industry. From the empirical evidence collected, it emerged that Microservices architectures are important in Software Engineering where 66.6% of the participants submitted that they use the Microservices Architecture to execute their daily duties. Microservices entail the design and implementation of software systems as a collection of small, independent services which perform specific functions. The Microservices architecture is gaining popularity in the Software Engineering domain in the sense that it provides scalability, flexibility, agility, and improved maintainability of products[28]. Despite numerous advantages offered by the Microservices Architecture, the technology is complex to implement[28]. Therefore, complexity can be simplified by introducing this technology into the Software Engineering curriculum. As stated by the Software Engineers in practice, this will surely simplify and equip the students with an understanding of the complex processes associated with this practice.

Containerisation and Orchestration were also ranked as important technologies in Software Engineering with 88.9% of the Software Engineers confirming that containers and the orchestration technology are very relevant in software development. Approximately 11.1% of the Software Engineers were not sure of the relevance of containerisation in Software Engineering. In general, containers are portable, scalable, consistent, and secure[29]. Moreover, containers enable organisations to accelerate software delivery while maintaining consistency and reliability across diverse environments. Technologies like Docker and Kubernetes are ideal for efficient deployment, scaling, and management of applications. This study established that containerisation and orchestration technologies are important topics in Software Engineering which should be incorporated into modern application development. Therefore, we can emphasize that students at tertiary institutions should be taught how to use containers in software development.

Out of the 36 Software Engineers who participated in this study, 88.9% considered Machine Learning (ML) and Artificial Intelligence (AI) as cutting-edge technologies. On the same premise, 5.6% of the Software Engineers were not sure of the relevance of ML and AI in Software Engineering. More so, 5.6% of the participants also confirmed that AI and ML is not relevant in their daily duties. Based on the study findings, we can settle on the conviction that ML and AI are important technologies to be embraced during software development. With specific emphasis on software development, the utilization of Artificial Intelligence and Machine Learning techniques for code analysis, testing, and automation is of paramount importance. It is therefore of paramount importance that Software Engineers in training apply the AI and ML principles in software development. It is vital to incorporate these technologies into the curriculum to keep students aligned with the industry expectations.

In Serverless Computing, software developers focus on writing and deploying individual pieces of code, without needing to manage the underlying infrastructure, such as servers, operating systems, or the runtime environments. Thus, this cloud computing model permits development and deployment of applications without managing server infrastructure directly. The study confirms that this is an important topic which is expected from IT graduates. It was interesting to note that 72.2% of the Software Engineers submitted that Serverless Computing is a very important concept in the Software Engineering domain. Only 27.8% of the participants were not sure of the relevance of serverless computing within the Software Engineering context. Serverless Computing provides a versatile, easily scalable, and developer-oriented method for creating and deploying cloud-native applications[30]. Understanding serverless computing is becoming more crucial in the industry as businesses embrace cloud-native structures and serverless solutions. When students learn about serverless computing at tertiary level, they acquire skills directly relevant to the job market, improving their employability and preparedness for contemporary software development positions.

Edge Computing aim to develop software solutions that leverage edge computing capabilities for faster processing and reduced latency. Incorporating edge computing into the software engineering curriculum is vital to provide students with the requisite knowledge, skills, and proficiencies required to navigate the dynamic realms of distributed computing. Such integration is poised to empower students, enabling them to contribute significantly to industry progress and innovation. 63.9% of the participants considered Edge Computing as an important topic in industry. On the same technology, 13.9% of the participants were non-aligned on the relevance of Edge Computing in Software Engineering. Lastly, 22.2% of the participants had negative views on the importance of Edge Computing within the Software Engineering context. Thus, Edge Computing is a very important topic which must be taught at undergraduate level.

The fact that Edge Computing was ranked low in comparison with other technologies in Software Engineering does not mean that the technology is irrelevant. The same technology should be embraced to produce graduates who are industry ready.

5. Conclusion and recommendations

Because Software Engineering is a dynamic field, there is need for continuously updating the curriculum. Based on empirical evidence from Software Engineering professionals in South Africa, some of the most important trends and technologies include the Microservices Architecture, Serverless Computing, Edge Computing and DevOps and Continuous Integration. The study therefore recommends that these technologies be integrated into the curriculum to empower students with the necessary skills expected from IT graduates. As highlighted by the Software Engineers in practice, understanding Serverless Computing is becoming more crucial in the industry as businesses embrace cloud-native structures and serverless solutions. When students learn about Serverless Computing at tertiary level, they acquire skills directly relevant to the job market, improving their employability and preparedness for contemporary software development positions.

Incorporating Edge Computing into the Software Engineering curriculum is vital to provide students with the requisite knowledge, skills, and proficiencies required to navigate the dynamic realms of distributed computing. The utilization of Artificial Intelligence and Machine Learning techniques for code analysis, testing, and automation is of paramount importance DevOps and continuous integration focus on automating the entire Software Development processes for increased efficiency and reliability. The study recommends that curriculum developers incorporate these technological trends into the Software Engineering curriculum. Further research should focus on the importance of these technologies in other domains.

Acknowledgements

The researcher would like to that all software engineers who participated in this study. Moreover, the researcher would like to thank the language editor who professionally edited this conference paper. May thanks are also accorded to all those who participated but are not listed here.

References

[1] N. Ouhbi, S. and Pombo, “Software engineering education: Challenges and perspectives,” *IEEE Glob. Eng. Educ. Conf.*, pp. 202–209, 2020, [Online]. Available: https://doi.org/10.1109/EDUCON45650.2020.9125353

[2] M. Towhidnejad, O. Ochoa, and A. Kiselev, “An Analysis of the Software Engineering Curriculum Using the Guideline Models,” *ASEE Southeast. Sect. Conf.*, 2020, [Online]. Available: https://sites.asee.org/se/wp-content/uploads/sites/56/2021/01/2020ASEESE83.pdf

[3] J. Annala, J. Lindén, M. Mäkinen, and J. Henriksson, “Understanding academic agency in curriculum change in higher education,” *Teach. High. Educ.*, vol. 28, no. 6, pp. 1310–1327, 2023, doi: 10.1080/13562517.2021.1881772.

[4] D. Mishra, A. and Mishra, “Sustainable Software Engineering: Curriculum Development Based on ACM/IEEE Guidelines,” *Softw. Sustain.*, pp. 269–285, 2021, [Online]. Available: https://doi.org/10.1007/978-3-030-69970-3_11

[5] Shamsuddeen abdullahi et al., “Software Testing: Review on Tools, Techniques and Challenges,” vol. 2, no. 2, pp. 11–18, 2020.

[6] M. Laplante, P.A. and Kassab, *What every engineer should know about software engineering*. Boca Raton: CRC Press, 2022.

[7] M. A. Ardis, “Curriculum Guidelines for Undergraduate Degree Programs in Software Engineering,” *IEEE Comput.*, vol. 48, no. 11, pp. 106–109, 2014, [Online]. Available: https://doi.ieeecomputersociety.org/10.1109/MC.2015.345

[8] M. K. Sharma, “A study of SDLC to develop well engineered software,” *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 3, 2017, [Online]. Available: http://www.ijarcs.info/

[9] S. Najihi, S. Elhadi, R. A. Abdelouahid, and A. Marzak, “Software Testing from an Agile and Traditional view,” *Procedia Comput. Sci.*, vol. 203, pp. 775–782, 2022, doi: 10.1016/j.procs.2022.07.116.

[10] M. A. Tian, F., Wang, T., Liang, P., Wang, C., Khan, A.A. and Babar, “The impact of traceability on software maintenance and evolution: A mapping study,” *J. Softw. Evol. Process*, vol. 33, no. 10, p. 2374, 2021, [Online]. Available: https://doi.org/10.1002/smr.2374

[11] S. Alsaqqa, S. Sawalha, and H. Abdel-Nabi, “Agile Software Development: Methodologies and Trends Blockchain technologies View project Social Media Networks View project,” *Artic. Int. J. Interact. Mob. Technol.*, pp. 246–270, 2020, [Online]. Available: https://doi.org/10.3991/ijim.v14i11.13269

[12] Z. Stojanov, “Software maintenance improvement in small software companies: Reflections on experiences,” *CEUR Workshop Proc.*, vol. 2913, pp. 182–197, 2021, doi: 10.47350/iccs-de.2021.14.

[13] B. Foster, E. and Towle Jr, *Software engineering: a methodical approach*. Auerbach Publications, 2021. [Online]. Available: https://doi.org/10.1201/9780367746025

[14] D. Benyon, *Designing User Experience - A Guide to HCI, UX And Interaction Design.*, 4th editio. Pearson Education, 2019.

[15] Divyani Shivkumar Taley, “Comprehensive Study of Software Testing Techniques and Strategies: A Review,” *Int. J. Eng. Res.*, vol. V9, no. 08, pp. 817–822, 2020, doi: 10.17577/ijertv9is080373.

[16] G. Agarwal, *Modern DevOps Practices: Implement and secure DevOps in the public cloud with cutting-edge tools, tips, tricks, and techniques*. Packet Publishing Ltd, 2021.

[17] K. Cao, Y. Liu, G. Meng, and Q. Sun, “An Overview on Edge Computing Research,” *IEEE Access*, vol. 8, pp. 85714–85728, 2020, doi: 10.1109/ACCESS.2020.2991734.

[18] J. Bogner, J. Fritzsch, S. Wagner, and A. Zimmermann, “Microservices in Industry: Insights into Technologies, Characteristics, and Software Quality,” *Proc. - 2019 IEEE*

Int. Conf. Softw. Archit. - Companion, ICSA-C 2019, no. February, pp. 187–195, 2019, doi: 10.1109/ICSA-C.2019.00041.

[19] P. Shafiei, H., Khonsari, A. and Mousavi, “Serverless computing: a survey of opportunities, challenges, and applications,” *ACM Comput. Surv.*, vol. 54, no. 11, pp. 1–32, 2022, [Online]. Available: https://doi.org/10.1145/3510611

[20] E. Casalicchio and S. Iannucci, “The state-of-the-art in container technologies: Application, orchestration and security,” *Concurr. Comput. Pract. Exp.*, vol. 32, no. 17, pp. 1–21, 2020, doi: 10.1002/cpe.5668.

[21] Q. Cao, K., Liu, Y., Meng, G. and Sun, “An overview on edge computing research.,” *IEEE Access*, no. 8, pp. 85714–85728, 2020, [Online]. Available: https://doi.org/10.1109/ACCESS.2020.2991734

[22] “Designing Survey Research: Recommendation for Questionnaire Development, Calculating Sample Size and Selecting Research Paradigms,” no. February, p. 2019, 2019.

[23] E. R. Babbie, *The practice of social research*. Cengage AU, 2020.

[24] IEEE Computer Society, *Software Engineering Competency Model Version 1.0*. 2014.

[25] B. Morais, P., Ferreira, M.J. and Veloso, “Improving student engagement with Project-Based Learning: A case study in Software Engineering,” *IEEE Rev. Iberoam. Tecnol. del Aprendiz.*, vol. 16, no. 1, pp. 21–28, 2021, [Online]. Available: https://doi.org/10.1109/RITA.2021.3052677

[26] P. Y. Hwang, G.J. and Chen, “Effects of a collective problem-solving promotion-based flipped classroom on students’ learning performances and interactive patterns.,” *Interact. Learn. Environ.*, vol. 31, no. 5, pp. 2513–2528, 2019, [Online]. Available: https://doi.org/10.1080/10494820.2019.1568263

[27] C. Jones, “A proposal for integrating DevOps into software engineering curricula,” in *In Software Engineering Aspects of Continuous Development and New Paradigms of Software Production and Deployment: First International Workshop, DEVOPS 2018, Chateau de Villebrumier, France, March 5-6, 2018*, 2019, pp. 33–47. [Online]. Available: https://doi.org/10.1007/978-3-030-06019-0_3

[28] J. Fritzsch *et al.*, “Adopting microservices and DevOps in the cyber-physical systems domain: A rapid review and case study,” *Softw. - Pract. Exp.*, vol. 53, no. 3, pp. 790–810, 2023, doi: 10.1002/spe.3169.

[29] A. Khan, “Key characteristics of a container orchestration platform to enable a modern application,” *IEEE Cloud Comput.*, vol. 4, no. 5, pp. 42–48, 2017, [Online]. Available: https://doi.org/10.1109/MCC.2017.4250933

[30] C. Papazov, Y., Sharkov, G., Koykov, G. and Todorova, “Managing Cyber-Education Environments with Serverless Computing,” *Digit. Transform. Cyber Secur. Resil. Mod. Soc.*, pp. 49–60, 2021, [Online]. Available: https://doi.org/10.1007/978-3-030-65722-2_4

An Accident-Avoidance Model for Driver Fatigue Detection using AIoT

Abstract. Many of all traffic accidents are attributed to drivers who are less vigilant. This leads to a number of fatalities on our roads While most drivers are aware of the risks associated with drinking and driving and texting while driving, many underestimate the hazards of driving when drowsy. There are very limited studied of fatigue detection using Artificial Intelligence of Things (AIoT).Systems for detecting driver weariness are being transformed by AIoT. According to literature, artificial intelligence systems use sophisticated algorithms and real-time data from IoT sensors to efficiently monitor driver behavior and spot signs of fatigue, lowering the likelihood of accidents on the road. The critical issue that a fatigue detection system must address is the question of how to detect fatigue accurately and early at the initial stage. This study explored the literature to find more on sensing and data collection, suitable machine learning algorithms for fatigue detection, and real-time monitoring to generate alerts.Finally the study developed a model and a flowchart for accident-avoidance.

Keywords: Driver safety, Artificial Intelligence of things, Internet of things, Fatigue and Accidents avoidance, Fatigue Detection

1 Introduction

The escalating incidence of traffic collisions resulting from a decline in driver alertness has emerged as a significant societal concern. Twenty percent of all traffic accidents are attributed to drivers who are less vigilant, according to statistics [1]. A study by the AAA Foundation for Traffic Safety revealed that tiredness was a contributing factor in 9.5% of all crashes and 10.8% of crashes involving airbag deployment, injury, or major property damage [2]. Data from the National Centre of Chronic Disease Prevention and Health Promotion indicates that individuals are three times more likely to be involved in a car accident while fatigued [2][3]. Additionally, 1 out of 25 drivers confessed to dozing off while driving [3]. While most drivers are aware of the risks associated with drinking and driving and texting while driving, many underestimate the hazards of driving when drowsy [4]. Moreover, collisions caused by driver hypo-vigilance are more severe than those caused by other types of accidents because drowsy drivers frequently fail to execute the appropriate precautions before a collision [5]. A driver's vigilance level pertains to their capacity to maintain a state of alertness and concentration while operating a motor vehicle [5]. It is intrinsically linked to driver weariness. As weariness develops, levels of vigilance diminish. Consequently, the driver's cognitive capacity to comprehend information, respond to potential dangers, and retain vehicle control declines [6], and thus, leading to falling asleep [7].

For this reason, developing systems for monitoring driver's level of vigilance and alerting the driver, when they are drowsy and not paying adequate attention to the road, is essential to reduce the number of accidents. The prevention of such accidents is a major focus of effort in the field of active safety research [8]. This research seeks to make a valuable contribution to the field by proposing an AIoT accident-avoidance model that aims to enhance the early detection of driver fatigue.

1.1 Artificial Intelligence of Things

The notion of Artificial Intelligence of Things combines the connection of the Internet of Things (IoT) with data-driven insights from Artificial Intelligence (AI) [9]. AIoT is transforming driver fatigue detection systems. AIoT systems utilize sophisticated algorithms and real-time data from IoT sensors to effectively monitor driver behaviour and identify indicators of weariness, hence reducing the risk of accidents on the road, [10] concurred. AIoT's precision in identifying driver drowsiness is unmatched, as it can assess several elements like eye motions, steering behaviors, and vehicle velocity to gauge the driver's state of attentiveness [11]. This proactive strategy not only boosts road safety but also enhances the driving experience for individuals and fleet operators.

One major benefit of AIoT is its real time nature [12], implying that the data sourced from IoT devices such as a smartphone or smartwatch, is analyzed on the spot, to provide accurate recommendations. This approach is famous for automated vehicles, traffic monitoring, and smart buildings [10][13][14]. Consequently, since early detection requires on-the-spot analysis, this concept is suitable for modelling an accident-avoidance model for detecting driver fatigue presented in this paper.

1.2 Aim of the study

The critical issue that a fatigue detection system must address is the question of how to detect fatigue accurately and early at the initial stage. Possible non-intrusive techniques for detecting fatigue in drivers using computer vision are methods based on the movement of the eyes and eyelids, methods based on head movements, and methods based on the opening of the mouth [8][11][12]. The researchers have selected this approach based on mouth opening and yawning in conjunction with driving behaviour at the time of suspected yawning, such as erratic steering, extracted from smartphone/smartwatch accelerometer and gyroscope sensors.

The paper is outlined as follows: Section 2 following Section 1 (Introduction) is presented next addressing the usage of AIoT in driver detection. In Section 3, a methodology to describe the process of model implementation is provided, highlighting the components of the Accident-Avoidance model. In Section 4, the model is evaluated, with case studies. Section 5 provides a conclusion and recommendations.

2 Driver Fatigue Detection

In this section, we examine the work around the use of AI to detect driver fatigue, and how these limitations can be addressed to improve results in terms of accuracy and performance. In investigating the literature, we ask the following questions:

- What are commonly used systems for detecting driver fatigue?
- What are the limitations of these commonly used systems?

The study by [15] evaluated the sensitivity of PERCLOS70 in detecting drowsiness levels to prevent crashes caused by drowsiness. The driving simulator was utilized to evaluate participants' performance in a drowsy state, despite its limitations in predicting increased crash risk. The simulator consisted of a facial camera, that will capture the driver's face and the data of wearable device (drowsimeter), which consisted of a high-speed camera, an infrared mirror, and an infrared light. The driving simulator effectively assessed driver performance and drowsy state, demonstrating the effectiveness of PERCLOS70 in identifying sleepiness changes. However, the authors have utilized an intrusive technique (the Drowsimeter), which is worn by the driver on the head.

In incorporating IoT in driver fatigue detection, [16] utilized eye-blink recognition technology, and vehicle sensors to detect signs of driver fatigue and prevent potential accidents. They further implemented a wearable carbon dioxide detection module to measure the air quality in a vehicle. Should the concentration be above 1500 ppm, they assume that fatigue should appear. Although they presented an implementation of drowsy driving prevention system, it shows the applicability of merging both AI and IoT to get improved results. However, it also makes use of vehicular sensors, which some vehicles may lack. The present accident-avoidance model employs smart device's sensors to attract a wider mass.

In eliminating intrusive wearables, [10] used video-based data from 13 drivers in real driving situations to create and test a method for detecting fatigue. They utilized a standard deep feature extraction module along with a customized sleepiness detection module to assess driver behaviour and facial expressions for indications of drowsiness. Although their results demonstrated promising outcomes, the study's limitation is its reliance on data from driving-based video games, which may not accurately represent real-world driving conditions. Using mobile device and sensor technology, [16] proposed a fatigue detection system based on Heart Rate Variability that collects data from a mobile device and sensor to observe changes in heart rate variability. This can detect patterns indicative of fatigue, which can be used for prediction purposes. Similarly, our paper proposes the use of a smart device, however, harnessing the power of having AI analytics on the spot, to analyze the data, and decide in real-time.

Significant progress has been made in developing technologies and methodologies to detect and prevent driver fatigue by studying the use of AI and IoT in Driver Fatigue Detection. Nevertheless, there are still gaps and areas that require additional attention and enhancement. This research establishes AIoT foundation for Driver Fatigue Detection, introducing a system that integrates multi-sensor data, overcomes intrusive detection limitations, and enables real-time monitoring.

3 Methodology

In this section, we explore the Accident-Avoidance Model, and the components that make it up. In depth, we will explore the sensing and data collection, suitable machine learning algorithms for fatigue detection, and real-time monitoring to generate alerts. In exploring the Accident-Avoidance model, we ask the following question:

- How can a driver fatigue detection system be implemented?

To answer this question, we present an architecture in figure 1 below suitable for AIoT projects. The architecture summarizes the steps that are listed hereafter.

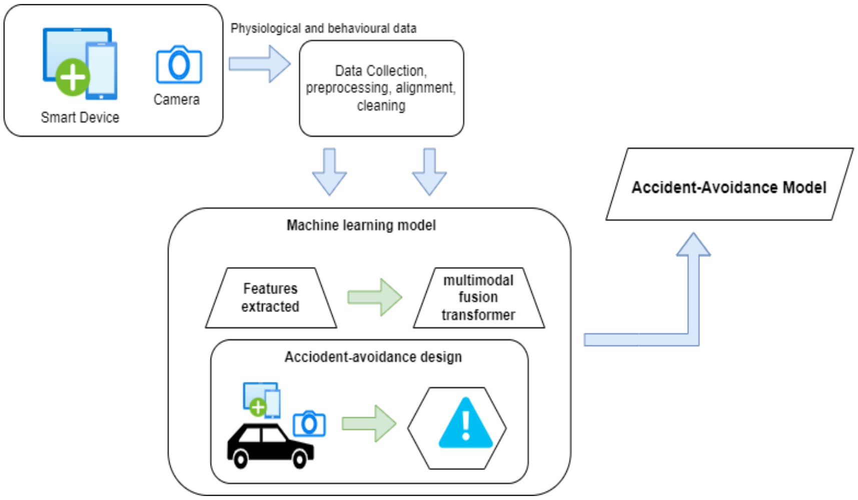

Smart Device
Camera
Physiological and behavioural data
Data Collection, preprocessing, alignment, cleaning
Accident-Avoidance Model
Machine learning model
Features extracted
multimodal fusion transformer
Acciodent-avoidance design

Fig. 1. Proposed Accident-Avoidance architecture

3.1 Proposed Accident-Avoidance Architecture Process Description

This sub-section outlines the overall process of the architecture depicted in figure 1. We delve into each step, providing insights on each method used and tools associated with it. The architecture functions over seven steps, namely data collection, feature extraction, machine learning model selection, fusion, real-time monitoring, alert mechanism, and validation and testing. These steps are discussed as follows.

Step 1: Data collection.

This study will use the Driver Drowsiness (DD) database to collect physiological data. Additionally, smart devices such as smartphones or smartwatches, and cameras will be used to collect physiological and motion data [18]. The DD database was downloaded from Kaggle. Table 1 categorizes collection of these data based on device type.

Table 1. Data collection per device type.

	Physiological data	**Motion data**
Smartphone / Smartwatch	Heart rate	Accelerometer
	Respiration rate	Gyroscope
	Eye tracking data	
Camera	Eyelid closure	
	Eye gaze	
	Pupil dilation	
	Facial expressions (yawns)	

Dealing with such data requires data anonymization as user privacy is crucial for protection. We used the data aggregation technique in which data summarization was applied. In this, raw data is converted to statistical summaries (means and ranges). Furthermore, we aggregated the data collected over short time windows to reduce the resolution using time series aggregation.

Step 2: Feature extraction.

From the real-time physiological and behavioral data collected, we extract relevant features, which are eye movement patterns, facial expressions, yawns, and vehicle parameters like sudden erratic steering. Table 2 categorizes these features by data type.

Table 2. Feature selection by data type.

	Physiological data
Heart rate	Average heart rate Heart rate variability
	Motion data
Accelerometer	Standard deviation of acceleration Frequency of sudden acceleration changes
Gyroscope	Frequency and duration of head nods Head tilt angle
	Camera Data
Eyelid closure	Blink rate Closed eye duration percentage
Eye gaze	Frequency of glances away from the road Duration of fixation on specific areas of interest
Pupil dilation	Average pupil diameter

Heart rate features

A significant increase or decrease from the driver's baseline will assist in determining fatigue or stress. Moreover, a reduced heart rate variability (HRV) can also suggest driver drowsiness.

Accelerometer and gyroscope features

A higher standard deviation of acceleration can indicate erratic movements, which are associated with drowsiness. Furthermore, sharp increases and decreases in speed can suggest compromised focus.

Excessive head nodding is another indicator of drowsiness. In addition, a sustained head tilt away from the road suggests inattentiveness.

Camera data features

In terms of eyelid closure, a significant increase in the blink rate can signal driver drowsiness. Using percentage of closed eye duration, we can quantify the severity of eyelid closure.

In terms of eye gaze, frequent glances away for extended periods could suggest distraction and drowsiness.

With regards to pupil dilation, identifying persistent dilation is crucial as it is linked to drowsiness and fatigue.

Step 3: Machine learning model

In this stage, we select machine learning models for real-time driver fatigue detection. We select support vector machine (SVM) based on the comparison from [21], where it achieves over 90%. Moreover, we opt for SVM as they are known for their efficiency in training and predictions, making them suitable for real-time applications. In improving the overall accuracy of driver fatigue detection, we combine SVM with CNN to achieve better eye tracking.

Step 4: Multi-modal fusion

The multi-modal early fusion approach will be implemented to get better and more accurate insights as data is from multiple heterogeneous channels across different modalities. We'll use a multimodal fusion transformer [17] because it is said to be better at classification tasks when compared to classical convolutional models. The objective is to enhance the accuracy and robustness of fatigue detection algorithms.

The early fusion approach works as follows:

- ***Data processing***: Raw data streams from the camera, and physiological sensors are combined prior to feeding them into the model.
- ***Feature engineering**:* Feature engineering techniques such as one-hot encoding are performed on the combined raw data to create a new feature set that leverages the strengths of both data sources.
- ***Feeding into SVM***: The combined feature set is then fed into the SVM for classification.

Step 5: Real-time monitoring

We then propose a design that serves as a concept for observing physiological and behavioral data to detect signs of fatigue during driving sessions.

Step 6: Alert mechanism

This feature forms as an extension to real-time monitoring, to quickly alert, warn, and notify the driver when signs of fatigue are detected.

Step 7: Validation and testing

Use cases will be to validate the accuracy of the accident-avoidance model proposed.

A flowchart of the major functions of the accident-avoidance model is shown in Figure 2 below. After obtaining the video file of the driver's image, it is converted into consecutive frames of images. The skin-color-based algorithm is applied to detect the face portion of the image. Since eyes lie in the upper half portion of the face, the lower half of the face is removed to narrow down the search area where the eyes exist. This reduces the amount of data in the image while retaining much of the critical information needed. The energy value of each frame is calculated, and it is used to differentiate between the open and closed eyes.

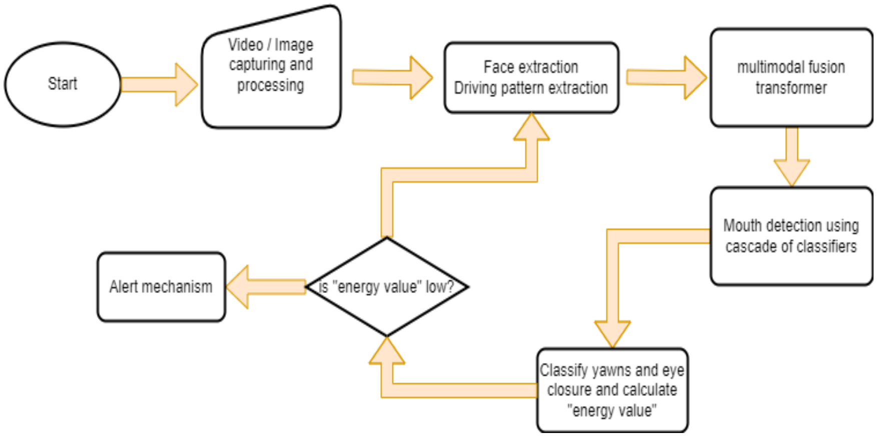

Start
Video / Image capturing and processing
Face extraction
Driving pattern extraction
multimodal fusion transformer
Mouth detection using cascade of classifiers
Classify yawns and eye closure and calculate "energy value"
is "energy value" low?
Alert mechanism

Fig. 2. Flowchart of the major functions of the accident-avoidance model.

3.2 Driver Drowsiness-Database:

The dataset includes recordings from four EEG channels, two EOG channels, and one ECG channel, along with annotation files for each volunteer and experiment trial. The signals and annotations are in edf format, generated twice for each volunteer, and include time marks for drowsiness, as recorded by the event push button. The labeling of files shows the information of #volunteer, gender, #trial, and signal type/channel.

The DD-Database is a collection of physiological signals from 10 healthy volunteers aged 20-50, collected during a driving simulator experiment. The data includes signals from four EEG channels, two EOG channels, and one ECG channel, along with annotation files. The experiment was conducted in two trials, each lasting 2 hours. The database contains 40 hours of information, contributing to the development and evaluation of algorithms for detecting drowsiness in drivers. The signals and annotations have labeling showing volunteer, gender, trial, and signal type/channel. Figure 3 below shows sample events that occurred.

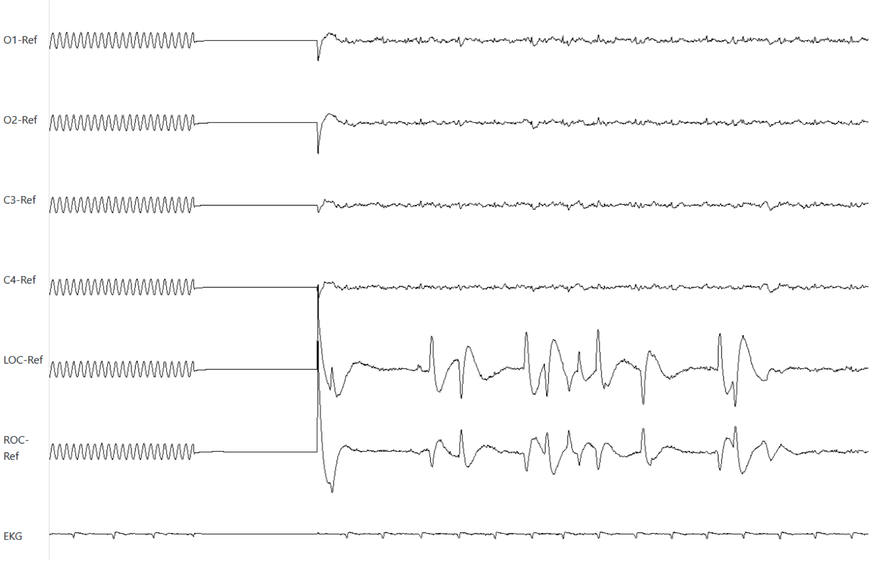


Fig. 3. Sample data of the DD-Database

3.3 Face detection:

Face location and detection are often the first steps in applications such as face recognition and/or facial expression analysis. In this paper, the detection of the face is performed using the following steps: Thresholding of an image is done from a grayscale image; it is used to create a binary image, which is used to distinguish between the object and the background. Thresholds are often determined based on surrounding lighting conditions and the complexion of the driver. To eliminate the noise left after thresholding, filtering is done by a Sobel filter. The Sobel Edge detector is used to detect edges based on applying a horizontal and vertical filter in sequence. The driver's face is detected using LBP (local binary pattern), which splits the image into four quadrants and then identifies the top and bottom portions."Fig.4" illustrates how LBP extracts a picture from a video, divides it into blocks, generates an LBP histogram from each block, and creates feature histograms.

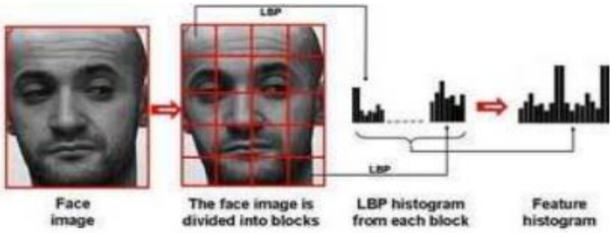

LBP
LBP
Face image
The face image is divided into blocks
LBP histogram from each block
Feature histogram

Fig. 4. Local Binary pattern

Below is the mathematical code used for face detection.

S= {∑, F, δ, C} S = Face Recognition. ∑ = set of input symbols = {Video File, image, character information} F = set of output symbol = {Match Found then notification to user, Not Found} δ = 1.

3.4 Mouth detection algorithm:

This mouth detection procedure classifies images based on the value of simple features. This image-based detection algorithm works on uncompressed images and has proven to be robust under various lighting conditions. The method is based on a cascade of boosted classifiers of simple Haar-wavelet-like features on different scales and positions. The features are brightness and contrast-invariant and consist of two or more rectangular region pixel sums that can be efficiently calculated by the canny integral image [19]. The feature set is overcomplete, and an adaptation of the AdaBoost learning algorithm is proposed to select and combine features into a linear classifier. To speed up detection, we used a cascade of classifiers such that every classifier could reject an image. All classifiers are trained to reject part of the candidates, such that on average only a small number of features are used per position and scale. After all possible mouth candidates are obtained, a grouping algorithm reduces groups of mouth candidates into single positive detections.

3.5 Fatigue detection

Fatigue detection has two phases: one is the training phase, and the other is the detection phase. SVM is a useful technique for data classification. A classification task usually involves training and testing data, which consists of some data instances. Each instance in the training set contains one “target value" (class labels) and “several attributes" (features). The goal of the SVM used in this paper is to produce a model that predicts the target value of data instances in the testing set, which are given only the attributes.

SVM requires that each data instance be represented as a vector of real numbers. Hence, if there are categorical attributes, we first must convert them into numeric data.

Scaling them before applying SVM is very important. The main advantage is to avoid having attributes in greater numeric ranges dominate those in smaller numeric ranges. The RBF kernel nonlinearly maps samples into a higher-dimensional space, so it, unlike the linear kernel, can handle the case when the relation between class labels and attributes is nonlinear. The second reason is the number of hyperparameters, which influences the complexity of model selection. The polynomial kernel has more hyperparameters than the RBF kernel. Finally, the RBF kernel has fewer numerical difficulties [20].

There are several advantages to SVMs. The most important advantage is that during the training process, only a few vectors out of the training set are selected to become support vectors. This reduces the computational cost and provides a better generalization. Another advantage is that there are no local minima in the quadratic program, so the solution found is always the optimum of the given training set. Finally, the main advantage is that the solution is not dependent on start conditions, unlike neural networks.

4 Discussions and Conclusion

This paper describes the various methods for detecting a driver's drowsiness by analyzing facial images taken by a camera installed in the dashboard. This system involves different steps such as Face detection, mouth detection fatigue detection then accident avoidance. The first step of face detection will be done through thresholding which will be performed so that the images may be clearly distinguished and do not include the unwanted background. The Sobel filter will then be used to eliminate the noise left after thresholding. In the second step, the AdaBoost learning algorithm will be used to combine features into a linear classifier to do mouth detection. Detection of the fatigue is done using SVM which is a useful technique for data classification. One of the challenges identified may be reading the photos after sunset may be difficult if there is inadequate light. This is in line with [21] who stated systems used in a controlled setting are a significant disadvantaged, since these systems must be evaluated in real-world driving situations under extremely safe settings before they can be considered reliable. This study also support [22] who combined CNN and VGG16 models, to detect facial sleepiness expressions and classified them into four categories (open, closed, yawning, and no yawning). The developed model may have trouble identifying the driver's eyeglasses wear. Future studies may include the use of light when it’s dark to do away with the darkness challenge.

References

1. Baikerikar, J., Kavathekar, V., John, R., Peeyus, L., Dharmadhikari, P. and Ghavate, N., 2021, December.Drowsiness Detection for Drivers. In 2021 International Conference on Advances in Computing, Communication, and Control (ICAC3) (pp. 1-5).IEEE.

2. Tefft, B.C.: Drowsy Driving in Fatal Crashes, United States, 2017-2021 (Research Brief). Washington, D.C.: AAA Foundation for Traffic Safety
3. Zoupos, A., Ζιακόπουλος, A., & Yannis, G. (2021). Modelling self-reported driver perspectives and fatigued driving via deep learning.: Traffic Safety Research, 1, 000003. https://doi.org/10.55329/galf7789
4. Schlick, C. J. R., Hewitt, D. B., Quinn, C., Ellis, R. J., Shapiro, K., Jones, A., … & Yang, A. D. (2020). A national survey of motor vehicle crashes among general surgery residents. Annals of Surgery, 274(6), 1001-1008. https://doi.org/10.1097/sla.0000000000003729
5. Greenlee, E. T., DeLucia, P. R., & Newton, D. C. (2024). Driver Vigilance Decrement is More Severe During Automated Driving than Manual Driving. Human Factors, 66(2), 574-588. https://doi.org/10.1177/00187208221103922
6. Hudson, A.N., Van Dongen, H.P.A. & Honn, K.A. Sleep deprivation, vigilant attention, and brain function: a review. *Neuropsychopharmacol.* **45**, 21–30 (2020). https://doi.org/10.1038/s41386-019-0432-6
7. Philip, P., Taillard, J., & Micoulaud-Franchi, J. (2019). Sleep Restriction, Sleep Hygiene, and Driving Safety: The Importance of Situational Sleepiness. Sleep medicine clinics, 14 4, 407-412
8. Magán, E., Sesmero, M. P., Alonso-Weber, J. M., & Sanchís, A. (2022). Driver drowsiness detection by applying deep learning techniques to sequences of images. Applied Sciences, 12(3), 1145. https://doi.org/10.3390/app12031145
9. Lin, Y., Chen, S., Tsai, C., & Lai, Y. (2021). Exploring computational thinking skills training through augmented reality and aiot learning. Frontiers in Psychology, 12. https://doi.org/10.3389/fpsyg.2021.640115
10. Bakker, B., Zabłocki, B., Baker, A. A., Riethmeister, V., Marx, B., Iyer, G., … & Ahlström, C. (2022). A multi-stage, multi-feature machine learning approach to detect driver sleepiness in naturalistic road driving conditions. IEEE Transactions on Intelligent Transportation Systems, 23(5), 4791-4800. https://doi.org/10.1109/tits.2021.3090272
11. Wang, Y., Liu, B., & Wang, H. (2022). Fatigue detection based on facial feature correction and fusion. Journal of Physics: Conference Series, 2183(1), 012022. https://doi.org/10.1088/1742-6596/2183/1/012022

12. Shahbakhti, M., Beiramvand, M., Nasiri, E., Far, S. M., Chen, W., Solé-Casals, J., … & Marozas, V. (2023). Fusion of eeg and eye blink analysis for detection of driver fatigue. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 31, 2037-2046. https://doi.org/10.1109/tnsre.2023.3267114
13. Sleem, A. and El-Henawy, I. M. (2023). Survey of artificial intelligence of things for smart buildings: a closer outlook. Journal of Intelligent Systems and Internet of Things, 8(2), 63-71. https://doi.org/10.54216/jisiot.080206
14. Zhang, J. and Tao, D. (2021). Empowering things with intelligence: a survey of the progress, challenges, and opportunities in artificial intelligence of things. IEEE Internet of Things Journal, 8(10), 7789-7817. https://doi.org/10.1109/jiot.2020.3039359
15. Murata, A., Doi, T., & Karwowski, W. (2022). Sensitivity of perclos70 to drowsiness level: effectiveness of perclos70 to prevent crashes caused by drowsiness. Ieee Access, 10, 70806-70814. https://doi.org/10.1109/access.2022.3187995
16. Santony, & Sani Muhamad Isa. (2022). Fatigue Detection System Based On Heart Rate Variability Using Mobile Device And Sensor. 17(11), 2349–2363. Https://Doi.Org/10.5281/Zenodo.7404555

17. Roy, S. K., Deria, A., Hong, D., Rasti, B., & Chanussot, J. (2023). Multimodal fusion transformer for remote sensing image classification. IEEE Transactions on Geoscience and Remote Sensing, 61, 1-20. https://doi.org/10.1109/tgrs.2023.3286826
18. Orosco, L., Garcés, M. A., Cañadas Fragapane, G. E., Dell'Aquila, C., Iturrieta Gimeno, J. C., & Laciar Leber, E. (2023). Drivers Drowsiness Database: A collection of physiological signals during the use of a driving simulator (DD-Database) [Data set]. Zenodo. https://doi.org/10.5061/dryad.5tb2rbp9c
19. Chen, S., Zhang, Y., Yin, B., & Wang, B. (2021). Trfh: towards real-time face detection and head pose estimation. Pattern Analysis and Applications, 24(4), 1745-1755. https://doi.org/10.1007/s10044-021-01026-3
20. Zhao, J., (2021). "The development and application of support vector machine." In *Journal of Physics: Conference Series*, vol. 1748, no. 5, p. 052006. IOP Publishing, 2021.
21. El-Nabi, S. A., El-Shafai, W., El-Rabaie, E. S. M., Ramadan, K. F., Abd El-Samie, F. E., & Mohsen, S. (2024). Machine learning and deep learning techniques for driver fatigue and drowsiness detection: a review. *Multimedia Tools and Applications*, *83*(3), 9441-9477.
22. Ahmed, M. I. B., Alabdulkarem, H., Alomair, F., Aldossary, D., Alahmari, M., Alhumaidan, M., ... & Zaman, G. (2023). A deep-learning approach to driver drowsiness detection. *Safety*, *9*(3), 65.

Blockchain in Healthcare: Implementing Hyperledger Fabric for Electronic Health Records at Frere Provincial Hospital

Olukayode Ayodele Oki[1][0000−0002−6887−9782],
Abayomi O. Agbeyangi[1][0000−0002−2504−2042], and
Aphelele Mgidi[1][0000−0002−1085−6465]

Walter Sisulu University, Buffalo City South Africa, {ooki,aagbeyangi}@wsu.ac.za

Abstract. As healthcare systems worldwide continue to grapple with the challenges of interoperability, data security, and accessibility, integrating emerging technologies becomes imperative. This paper investigates the implementation of blockchain technology, specifically Hyperledger Fabric, for Electronic Health Records (EHR) management at Frere Hospital in the Eastern Cape province of South Africa. The paper examines the benefits and challenges of integrating blockchain into healthcare information systems. Hyperledger Fabric's modular architecture is harnessed to create a secure, transparent, and decentralized platform for storing, managing, and sharing EHRs among stakeholders. The study used a mixed-methods approach, integrating case studies and data collection methods through observation and informal questions, with the specific goal of understanding current record management methods and challenges. This method offers practical insights and validates the approach. The result demonstrates the role of blockchain in transforming healthcare, framed within a rigorous exploration and analysis. The findings of this study have broader implications for healthcare institutions seeking advanced solutions to address the persistent challenges in electronic health record management. Ultimately, the research underscores the transformative potential of blockchain technology in healthcare settings, fostering trust, security, and efficiency in the management of sensitive patient data.

Keywords: Blockchain · E-Health · Hyperledger Fabric · Electronic Health Records.

1 Introduction

Blockchain technology has expanded beyond its original connection with financial technology and is now making its way into different areas, such as healthcare [1, 2]. Blockchain is recognised for its ability to disrupt banking institutions and its close association with cryptocurrencies. It is now being seen as a promising way to improve data security, transparency, and interoperability in healthcare environments [3]. Initially, cryptocurrencies like Bitcoin relied on blockchain as their foundational technology. It functions as a decentralised and unchangeable system of record-keeping. By utilising cryptographic techniques, it ensures transparency, security, and tamper-resistance when storing transactions or data records across a dispersed network of nodes. According to Tseng and Shang [4], it removes the necessity for intermediaries seen in conventional centralised databases, enabling direct transactions between peers while preserving the integrity of data.

Managing electronic health records (EHR) in the traditional healthcare system encounters notable hurdles such as data security risks, interoperability problems, and administrative inefficiencies. The problems impact patient privacy, impede smooth information sharing between healthcare professionals, and lead to inefficiencies in healthcare service delivery. Despite attempts to tackle these issues with technical progress, current solutions have not fully achieved widespread and lasting enhancements in EHR management. Hyperledger Fabric [5], an open-source blockchain platform designed for enterprise use, shows great potential for transforming multiple industries, such as healthcare. Its distinctive attributes make it especially suitable for tackling the issues found in conventional healthcare information management systems [6]. By providing unique benefits compared to other blockchain frameworks and conventional data management methods, it is suitable in healthcare settings. For example, Hyperledger Fabric's permissioned blockchain network allows for greater control over who can access and contribute to the stored data, ensuring privacy and security. By harnessing the power of this innovative technology, healthcare facilities can enhance the quality of care they provide while reducing costs and increasing patient trust.

The motivation for the study is due to the benefits of this technological innovation as well as its security, technical aspects, and operational feasibility at healthcare facilities. The current increase in specialised healthcare services and patient movement emphasises the crucial need to provide healthcare facilities with thorough patient medical histories. Sharing clinical data securely across several healthcare facilities is a complex task that demands a careful equilibrium between confidentiality, data integrity, and patient privacy.

The main contribution of the paper is to explore the transformative potential of implementing a blockchain-based Electronic Health Record (EHR) system. aiming to revolutionise healthcare data management and improve patient care outcomes. Other specific contributions of the paper include:

- examining the viability, advantages, and obstacles of implementing a blockchain-based EHR system
- the implementation of the blockchain-based EHR system.

The Frere Hospital, where the study is conducted as a case study, is a provincial and public hospital, and it plays a vital role as the primary medical facility catering to the healthcare needs of East London and the broader Eastern Cape region in South Africa. It focuses on providing essential medical services to underprivileged communities, ensuring access to quality healthcare for all. By introducing a blockchain-based solution, the study seeks to demonstrate the potential benefits of blockchain technology in improving data security, integrity, and accessibility in healthcare settings. While the system is initially an ongoing research project, its ultimate goal is to showcase its feasibility and effectiveness in real-world healthcare environments and potentially pave the way for adoption.

The remaining part of the paper is structured as follows: Section 2 discusses the literature review; Section 3 explains the methods; the discussion on the implementation is in Section 4; and Section 5 concludes the paper.

2 Literature review

2.1 Background

Blockchain has emerged as a very significant and rapidly evolving subject on a global scale, especially in the context of financial technologies, since the start of the 21st century. This technology, in conjunction with distributed database technologies, plays a crucial role in facilitating advancements in distributed transaction and ledger systems, hence creating fresh possibilities for digital platforms and services [14, 13]. According to Bhutta et al. [15], the blockchain environment consists of a decentralised system that uses cryptography to record and store an unchangeable, reliable, and sequential log of transactions amongst interconnected participants. Similar to a distributed ledger, the parties involved maintain, update, and validate this system. This approach eliminates the necessity of a central authority to authenticate transactions, as all members of the network collectively verify and safeguard the information [16]. The extensive implementation of blockchain technology across multiple sectors, such as finance[17], supply chain management[14], healthcare[2], and voting systems[18], is mostly due to its high level of transparency and security. The potential uses of blockchain technology are extensive, with certain experts forecasting that it has the ability to fundamentally transform the methods by which we engage in commerce and transfer assets in the coming years.

The healthcare management sector is constantly changing as new technical advancements provide fresh ways to tackle long-standing challenges. Blockchain technology's implementation in healthcare represents a fundamental shift in the management of electronic health records (EHRs) and patient data [3]. Although traditional EHR systems have been effective in converting medical records into digital format and improving administrative efficiency, they frequently encounter challenges with the accuracy and security of data, as well as the capacity to seamlessly exchange information with other systems. Menachemi and Collum (2011) opined that EHR systems face some challenges, including high upfront costs, ongoing maintenance expenses, workflow disruptions leading to temporary productivity losses, and potential privacy concerns among patients. Specifically, blockchain provides a secure and verifiable record of patient data exchanges, which helps address these difficulties. It enhances data security and confidentiality by decentralising data storage and employing cryptographic techniques to prevent unauthorised access, data breaches, and manipulation.

Furthermore, the distributed architecture of blockchain allows for effortless compatibility between different healthcare systems and participants, facilitating the safe transfer of patient information between different organizations. Interoperability is especially vital in healthcare settings, as patient care frequently requires cooperation across many clinicians and institutions. The lack of interoperability, as stated by Iroju et al. [20], poses significant challenges in healthcare systems, where health practitioners may face difficulties in obtaining complete patient information, leading to repeated tests and procedures. Healthcare organisations may optimise data sharing processes, boost care coordination, and improve patient outcomes by utilising blockchain technology. Also, blockchain technology has the potential to empower patients by giving them more authority over their health data, enabling consent management, and promoting transparency in healthcare transactions [21].

Although blockchain has great promise, its implementation in the healthcare industry is not without obstacles. According to Sharma and Joshi [22], low awareness of legal issues and inadequate support from high-level management are the most influential driving barriers, based on their study on blockchain adoption in the Indian healthcare industry. However, the investigation of blockchain technology in healthcare gives a promising chance to tackle the intricacies of EHR management, boost data security and interoperability, and eventually improve the provision of healthcare services to patients.

2.2 Related Work

In healthcare, studies have been conducted on the impact and feasibility of blockchain to enhance healthcare service delivery [6–11, 21, 22]. Several of these studies have shown promising results, particularly in enhancing patient data security and improving interoperability among different healthcare providers. It was stated by Pandey and Litoriya [7] that blockchain provides a secure and transparent system for integrated healthcare services, ensuring corruption-free and efficient implementation of nationwide health insurance programs. Furthermore, research has also highlighted the potential of blockchain technology in reducing medical errors [9, 8] and improving overall patient care quality. Overall, the existing literature emphasises the importance of analysing and designing blockchain-based EHR systems to revolutionise healthcare delivery. An important feature of blockchain technology in healthcare, specifically using Hyperledger Fabric, is its capacity to guarantee the integrity and confidentiality of data [10]. This, in turn, improves patient privacy and security. In the study by Sutradhar et al. [6], the paper presents a specialised framework for managing identity and access in the healthcare industry using blockchain technology. By utilising Hyperledger Fabric and OAuth 2.0, it was shown that the framework guarantees improved security and scalability while also providing transparency, immutability, and fraud prevention.

Another interesting approach was by Antwi et al. [10]; the study examines the practicality of using private blockchain technologies, notably Hyperledger Fabric, to address different requirements and scenarios in the healthcare industry. Empirical assessments in the study reveal the significant advantages of the approach, such as heightened security, adherence to regulations, interoperability, adaptability, and scalability, hence demonstrating their potential to efficiently tackle crucial issues in healthcare data management. The study by Stamatellis et al. [11] also looked at the changes that have happened in healthcare from using paper-based systems to electronic health records (EHRs) and the problems that have come with them, mainly looking at cybersecurity risks like malware and ransomware. It specifically focuses on notable events like the WannaCry ransomware cryptoworm and the Medjack cyberattack. The research highlights the susceptibility of conventional EHR databases to typical methods of attack and emphasises the necessity for a scalable, unchangeable, transparent, and secure solution to tackle these difficulties. The results of the experimental evaluation provide evidence of the advantages of the method in improving security, compliance, compatibility, flexibility, and scalability. This offers a potential alternative for managing decentralised medical data, as also implied in this research.

Similarly, an effort by Al-Sumaidaee et al. [12] addresses the difficulties in the healthcare sector caused by inadequate interoperability and fragmented communication systems, which have negative effects on patients, resources, and costs. The paper suggests implementing blockchain technology, specifically Hyperledger Fabric, to enhance the exchange of information among various healthcare organisations by introducing a trust element. It emphasises the substantial influence of factors such as workforce size and transactions per second (TPS) on network latency and overall throughput. The studies emphasise the potential of blockchain technology, specifically Hyperledger Fabric, in strengthening healthcare service delivery, improving patient data security, minimising medical errors, and transforming healthcare practices. Although recent studies show encouraging outcomes, there are still unresolved concerns that need to be tackled. Some of the challenges include scalability, regulatory compliance, interoperability, and the requirement for additional empirical validation in real-world healthcare environments. This study conducted empirical assessments and real-world implementations to provide practical insights and validation for the solutions, considering these issues. It is aimed at enhancing patient outcomes and healthcare service delivery by raising awareness and acceptance of blockchain technology in the healthcare sector.

3 Methods

The study uses a mixed-methods approach, starting with requirement gathering, to investigate the adoption of blockchain technology, specifically the Hyperledger Fabric, for the management of electronic health records (EHR) at Frere Provincial Hospital. The implementation was divide into separate components, each of which fulfils a specific role in the system's functionality. The main constituents consist of the Hyperledger Fabric network and distributed ledger, smart contracts, the application Software Development Kit (SDK), and the application frontend.

3.1 Requirement Gathering

The research collected data as part of the requirement gathering for the system implementation. The goal was to gather patient insights while also observing the operational dynamics at Frere Provincial Hospital. A structured questionnaire was designed as a form of standardised format for data collection. To encourage open and honest feedback, the respondents were assured of the anonymity of their responses. It was distributed to a sample group of patients, encompassing a diverse demographic. This was to investigate concerns related to the current non-digital storage of medical records and assess the perceived impact of administrative processes on the overall service delivery experience.

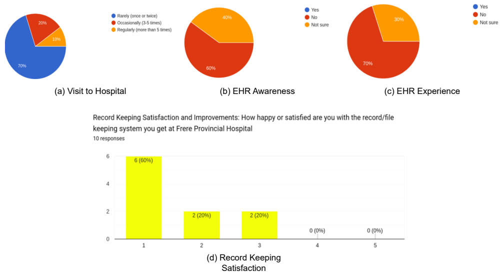
For the on-site operational observations, physical observation was conducted in the hospital environment, paying attention to document handling, data retrieval, and administrative workflows. This was followed by an informal interview with the hospital staff to gather qualitative insights into the challenges faced in data management. Prior informed consent was obtained from participants in adherence to ethical considerations without collecting personally identifiable information.

Figure 1 provides a visual representation of the requirement-gathering analysis, based on responses from a random sample of 10 patients. Subfigure 1(a) shows the frequency of patient visits to the hospital, with 70% of respondents visiting regularly (more than five times), indicating a high reliance on the hospital's services. Subfigure 1(b) illustrates the awareness of electronic health records (EHR) among patients, revealing that 60% of respondents are not aware of EHRs, suggesting a significant gap in patient knowledge about digital health record systems. Subfigure 1(c) depicts the experience of patients with EHRs; 70% of respondents have no prior experience with EHRs, highlighting the lack of exposure and potential challenges in transitioning from a non-digital to a digital record-keeping system. Finally, subfigure 1(d) reveals that 60% of respondents are dissatisfied with the current manual record-keeping system, pointing to a critical need for improvement in the hospital's data management processes. This data underscores the necessity for implementing a blockchain-based EHR system to enhance efficiency, security, and patient satisfaction.

3.2 The Hyperledger Fabric Components

The Hyperledger Fabric's components are crucial and provide the technical foundation for understanding the application of the proposed blockchain technology implementation for electronic health records (EHR).

The modular architecture of the Hyperledger Fabric, as described in Figure 2, is specifically designed to offer a flexible and scalable framework for blockchain solutions, forming the backbones of the entire application, managing the ledger, and facilitating interactions between various participants. The smart contracts are crafted using JavaScript, a versatile and widely used programming language. These contracts define the transactional logic and rules governing the interactions between different entities within the Hyperledger Fabric network. The front



(a) Visit to Hospital (b) EHR Awareness (c) EHR Experience

(d) Record Keeping Satisfaction

Fig. 1. Insights on Electronic Health Record System: (a) Frequency of hospital visits among patients (b) Awareness of Electronic Health Records (EHR) (c) Experience with using EHR systems (d) Current record keeping satisfaction.

end of the application used Angular, a popular web application framework. Angular[1] provides a robust and dynamic user interface, enabling a user-friendly experience. It serves as the interface through which users (patients, doctors, and administrators) interact with the system. The test network adaptation entails transforming organizations within the network to represent distinct hospitals. This modification entails adjusting Docker files, configuration files, and corresponding certificates. Specifically, alterations are made to organization names in the *"configtx.yaml"* file, and adjustments are applied to the certificate authorities in the docker-compose file. Afterwards, the files that reference these configurations are updated. Once this custom network is set up and a designated channel is established, as shown in the architecture, the necessary certificates are made for the organisation and peers that are involved. Also, as illustrated in Figure 2, the Hyperledger Fabric architecture comprises two main organisations: Org1 (hospitals or clinics) and Org2 (patients). Each organisation has peers—Peer0 for doctors and Peer1 for patients—connected via Channel 1, which ensures secure, transparent, and verifiable data exchanges. Both peers maintain a ledger (L1) that records all transactions. The Client SDK facilitates application interactions with the blockchain network, while the Membership Service Provider (MSP) handles identity management, ensuring authorised access. The Certificate Authority (CA) issues digital certificates for identity validation, enhancing the system's security. This architecture supports secure, efficient data sharing, giving patients control over their health records, and enabling healthcare providers to update patient information as needed.

3.3 Smart contract implementation and role-based access control

In line with the application's modularity, all components are pluggable. The backend code and smart contracts utilise *ExpressJS* as a server to deliver REST

[1] https://angular.io/

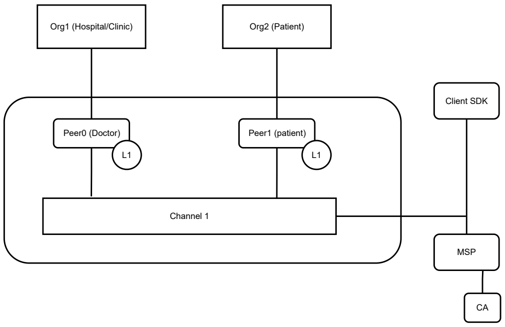


Fig. 2. Overview of the Hyperledger Fabric Architecture

APIs. Communication between the frontend and backend is facilitated through REST calls, incorporating JSON web tokens for authentication. he fabric framework supports two databases - *LevelDB* and *CouchDB*. The *CouchDB* database was chosen due to its enhanced flexibility, especially in handling images and supporting indexes, and features vital for the healthcare context. Given that all patient data resides within the *CouchDB*, obviating the need for a separate Electronic Health Record (EHR) store, *CouchDB* aligns perfectly with the system's requirements. In this architecture, *CouchDB* is leveraged to store the world state. This design choice eliminates the need to query the entire transaction log for each transaction request, streamlining system efficiency. By providing quicker access to relevant information without traversing the entire transaction history, *CouchDB* optimises the overall performance of the system.

In the application, all the executable business logic is encapsulated within smart contracts. This implies that any operations involving the creation, retrieval, update, or deletion of records in the distributed ledger are orchestrated through smart contracts. Each functionality required by the system to interact with the Hyperledger Fabric (HLF) network is encapsulated within a separate function for clarity and ease of maintenance.

As depicted in Figure 3, there are three primary smart contracts packaged into a single *chaincode*, each catering to specific roles (Admin, Patient, Doctor):

- AdminContract: Invoked by the administrator, this contract endows the admin with the capability to create and delete patient records by adding or removing patient objects from the ledger. The admin can also retrieve information on all patients across the network.
- PatientContract: Designed for patient interactions with the ledger, this contract encapsulates logic tailored for the patient's role. Specifically, the patient can update and view personal details and passwords through the methods defined in the contract. Additionally, the patient contract includes methods to grant and revoke access from a doctor.

- DoctorContract: Tailored for doctor interactions, this contract provides methods enabling doctors to update and read patients' medical details.

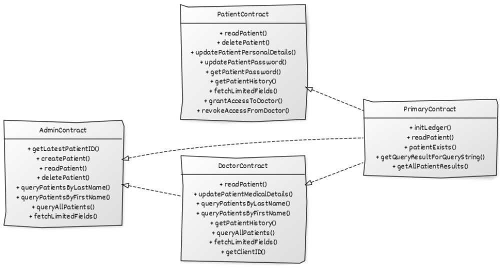


Fig. 3. Relationships and Dependencies between the Contract Classes

The main reason for the three smart contracts is to ensure that each role has appropriate access to data on the ledger. Only the *PatientContract* has the right to update personal details, grant or revoke access, and update the password. Similarly, the doctor's methods are inaccessible to the patient or admin. While the *readPatient* method is common for all roles, the data retrieved by the contracts varies based on the role (i.e., the admin has access to patient names, the doctor has access to medical records, and the patient has access to the entire patient objects). This is to ensure a granular and role-specific access control mechanism within the system on the Hyperledger Fabric. The breakdown of the key components of the hyperledger fabric in Algorithm 1 (Figure 4) illustrates the steps involve in setting up the network, deploying smart contracts, and performing transactions.

4 Discussion

The implementation of the blockchain-based Electronic Health Records (EHR) management system at Frere Provincial Hospital yielded promising results across various test cases involving three sets of users: administrators, doctors, and patients. Through rigorous testing, the system demonstrated its efficacy in enhancing data security, streamlining administrative processes, and improving patient care delivery. Figure 5 shows a sample deployment of a local test fabric network, the installation and instantiation of a healthcare chaincode, and the visualisation of the transaction history using the Hyperledger Explorer.

The testing was conducted in separate phases, specifically focusing on assessing the specific features designed for administrators, doctors, and patients. Every stage yielded valuable observations regarding the system's performance,

Algorithm 1 Setting Up Hyperledger Fabric Network and Smart Contracts

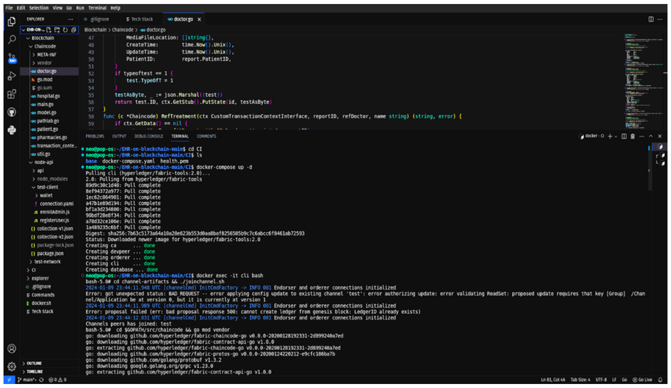
```
Initialization
Define Org1 (Hospital/Clinic) and Org2 (Patients)
Create Peer0 for Org1 and Peer1 for Org2
Establish Channel 1 for communication between peers
Configure Membership Service Provider (MSP) for identity management
Setup Certificate Authority (CA) for digital certificates issuance
Smart Contracts Development
Develop AdminContract, PatientContract, and DoctorContract using JavaScript
Package contracts into chaincode
Network Deployment
Deploy chaincode onto the Hyperledger Fabric network
Initialize CouchDB for world state storage
Transactions Management
AdminContract
Create and delete patient records
Retrieve information on all patients
PatientContract
Update and view personal details and passwords
Grant and revoke access to doctors
DoctorContract
Update and read patients' medical details
Access Control
Ensure role-specific access to data on the ledger
Implement granular access control for Admin, Patient, and Doctor roles
```

Fig. 4. Algorithm for Setting Up Hyperledger Fabric Network and Smart Contracts

usability, and impact on healthcare service delivery. Figure 6 shows the complete Hyperledger Explorer integration.

Overall, our findings from comparing traditional record systems, analysing data from requirements gathering, and implementing the blockchain-based electronic health records (EHR) systems reveal the following:

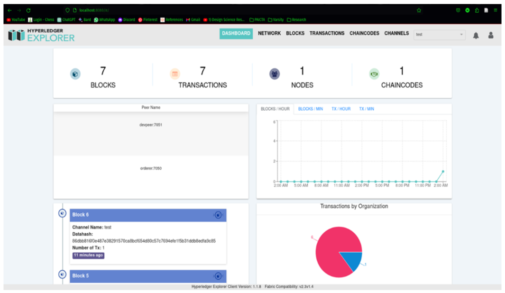
- Hyperledger Fabric's decentralised structure eliminates vulnerabilities associated with centralised traditional database systems.
- The automatic correction mechanisms of Hyperledger Fabric enhance its effectiveness in mitigating manipulation risks. This provides reassurance regarding the integrity and authenticity of healthcare data, addressing concerns prevalent in traditional database systems.
- Hyperledger Fabric's ledgers ensure data permanence, contrasting with the mutable nature of traditional databases. This highlights the significance of



```go
            MediaFileLocation: []string{},
            CreateTime:        time.Now().Unix(),
            UpdateTime:        time.Now().Unix(),
            PatientID:         report.PatientID,
        }
        if typeoftest == 1 {
            test.TypeOfT = 1
        }
        testAsByte, _ := json.Marshal((test))
        return test.ID, ctx.GetStub().PutState(id, testAsByte)
}
func (c *Chaincode) RefTreatment(ctx CustomTransactionContextInterface, reportID, refDoctor, name string) (string, error) {
        if ctx.GetData() == nil {
```

```
neo@pop-os:~/EHR-on-blockchain-main$ cd CI
neo@pop-os:~/EHR-on-blockchain-main/CI$ ls
base  docker-compose.yaml  health.pem
neo@pop-os:~/EHR-on-blockchain-main/CI$ docker-compose up -d
Pulling cli (hyperledger/fabric-tools:2.0)...
2.0: Pulling from hyperledger/fabric-tools
89d9c30c1d48: Pull complete
8ef94372a977: Pull complete
1ec62c064901: Pull complete
a47b1e89d194: Pull complete
bf1a3d234800: Pull complete
90bdf28e8f34: Pull complete
a78d32ce106e: Pull complete
1a489235c6bf: Pull complete
Digest: sha256:7b63c5173a64a10a28e623b553d0aa8baf8256505b9c7c6abcc6f8461ab72593
Status: Downloaded newer image for hyperledger/fabric-tools:2.0
Creating ca       ... done
Creating devpeer  ... done
Creating orderer  ... done
Creating cli      ... done
Creating database ... done
neo@pop-os:~/EHR-on-blockchain-main/CI$ docker exec -it cli bash
bash-5.0# cd channel-artifacts && ./joinchannel.sh
2024-01-09 23:44:11.948 UTC [channelCmd] InitCmdFactory -> INFO 001 Endorser and orderer connections initialized
Error: got unexpected status: BAD_REQUEST -- error applying config update to existing channel 'test': error authorizing update: error validating ReadSet: proposed update requires that key [Group]  /Chan
nel/Application be at version 0, but it is currently at version 1
2024-01-09 23:44:11.989 UTC [channelCmd] InitCmdFactory -> INFO 001 Endorser and orderer connections initialized
Error: proposal failed (err: bad proposal response 500: cannot create ledger from genesis block: LedgerID already exists)
2024-01-09 23:44:12.031 UTC [channelCmd] InitCmdFactory -> INFO 001 Endorser and orderer connections initialized
Channels peers has joined: test
bash-5.0#  cd $GOPATH/src/chaincode && go mod vendor
go: downloading github.com/hyperledger/fabric-chaincode-go v0.0.0-20200128192331-2d899240a7ed
go: downloading github.com/hyperledger/fabric-contract-api-go v1.0.0
go: extracting github.com/hyperledger/fabric-chaincode-go v0.0.0-20200128192331-2d899240a7ed
go: downloading github.com/hyperledger/fabric-protos-go v0.0.0-20200124220212-e9cfc186ba7b
go: downloading github.com/golang/protobuf v1.3.2
go: downloading google.golang.org/grpc v1.23.0
go: extracting github.com/hyperledger/fabric-contract-api-go v1.0.0
```

Fig. 5. The Fabric Network Deployment



Channel Name: test
Datahash: 86dbb816f0e487e38291570ca8bcf654d80c57c7694efe1f5b31ddb8edfa9c85
Number of Tx: 1

Fig. 6. The Hyperledger Explorer Integration

data integrity and permanence in healthcare settings, where accurate and reliable records are paramount.

5 Conclusion

The study examines the implementation of Hyperledger Fabric, a type of blockchain technology, for managing electronic health records at Frere Provincial Hospital. The study demonstrated the pragmatic application and advantages of blockchain technology in healthcare information systems. The study offers insightful information about how blockchain technology can revolutionise healthcare information systems by enhancing EHR management efficiency, security, and privacy. Although our study highlights the potential use of blockchain technology in the management of electronic health records, it is important to acknowledge our research's limitations. As part of our ongoing exploration of this technology, the study is a preliminary investigation into the practical deployment and advantages of blockchain technology in healthcare settings. The findings, derived from a single case study, may not comprehensively encompass the intricacies and complexities of wider healthcare settings. In light of our practical implementation of the system, additional investigation and experimentation are necessary to thoroughly examine the complete functionalities and prospective uses of blockchain technology in healthcare settings. Specifically, challenges persist within the Hyperledger Fabric necessitating ongoing development, such as transaction history retrieval issues and performance concerns, which underscore the need for continued refinement.

Acknowledgments

The authors extend sincere gratitude to Frere Provincial Hospital for their collaboration and support during the implementation of the blockchain-based Electronic Health Records (EHR) system. The opinions expressed in this study are solely those of the research team.

References

1. Khezr, S., Moniruzzaman, M., Yassine, A., Benlamri R.: Blockchain Technology in Healthcare: A Comprehensive Review and Directions for Future Research. Applied Sciences **9**(9), 1736 (2019). https://doi.org/10.3390/app9091736
2. Massaro, M.: Digital transformation in the healthcare sector through blockchain technology. Insights from academic research and business developments. Technovation, 120, 102386, (2023).
3. Yaqoob, I., Salah, K., Jayaraman, R., Al-Hammadi, Y.: Blockchain for Healthcare Data Management: Opportunities, Challenges, and Future Recommendations. Neural Computing and Applications. **34**, 11475–11490 (2022). https://doi.org/10.1007/s00521-020-05519-w
4. Tseng, C.-T.; Shang, S.S.C.: Exploring the Sustainability of the Intermediary Role in Blockchain. Sustainability. **13**, 1936 (2021). https://doi.org/10.3390/su13041936
5. Androulaki E, Barger A, Bortnikov V, Cachin C, Christidis K, De Caro A, Enyeart D, Ferris C, Laventman G, Manevich Y, Muralidharan S.: Hyperledger fabric: a distributed operating system for permissioned blockchains. In Proceedings of the Thirteenth EuroSys Conference (EuroSys '18). Association for Computing Machinery, New York, NY, USA, Article 30, 1–15 (2018). https://doi.org/10.1145/3190508.3190538
6. Sutradhar S, Karforma S, Bose R, Roy S, Djebali S, Bhattacharyya D.: Enhancing identity and access management using Hyperledger Fabric and OAuth 2.0: A block-chain-based approach for security and scalability for healthcare industry. Internet of Things and Cyber-Physical Systems. 4:49-67 (2024). https://doi.org/10.1016/j.iotcps.2023.07.004
7. Pandey P, Litoriya R.: Implementing healthcare services on a large scale: challenges and remedies based on blockchain technology. Health Policy and Technology. 9(1):69-78 (2020). https://doi.org/10.1016/j.hlpt.2020.01.004

8. Zarour M, Ansari MT, Alenezi M, Sarkar AK, Faizan M, Agrawal A, Kumar R, Khan RA.: Evaluating the impact of blockchain models for secure and trustworthy electronic healthcare records. IEEE Access. 8:157959-73 (2020). https://doi.org/10.1109/ACCESS.2020.3019829
9. Radanović I, Likić R. Opportunities for use of blockchain technology in medicine. Applied health economics and health policy. 16:583-90 (2018). https://doi.org/10.1007/s40258-018-0412-8
10. Antwi M, Adnane A, Ahmad F, Hussain R, ur Rehman MH, Kerrache CA.: The case of HyperLedger Fabric as a blockchain solution for healthcare applications. Blockchain: Research and Applications. 2(1):100012 (2021). https://doi.org/10.1016/j.bcra.2021.100012
11. Stamatellis C, Papadopoulos P, Pitropakis N, Katsikas S, Buchanan WJ. A privacy-preserving healthcare framework using hyperledger fabric. Sensors. 2020 Nov 18;20(22):6587. https://doi.org/10.3390/s20226587
12. Al-Sumaidaee G, Alkhudary R, Zilic Z, Swidan A. Performance analysis of a private blockchain network built on Hyperledger Fabric for healthcare. Information Processing & Management. 60(2):103160 (2023). https://doi.org/10.1016/j.ipm.2022.103160
13. Gad, Ahmed G., Diana T. Mosa, Laith Abualigah, and Amr A. Abohany. "Emerging Trends in Blockchain Technology and Applications: A Review and Outlook." Journal of King Saud University - Computer and Information Sciences 34, no. 9 (2022): 6719–42. https://doi.org/10.1016/j.jksuci.2022.03.007.
14. Dutta, Pankaj, Tsan-Ming Choi, Surabhi Somani, and Richa Butala. "Blockchain Technology in Supply Chain Operations: Applications, Challenges and Research Opportunities." Transportation Research Part E: Logistics and Transportation Review 142 (2020): 102067. https://doi.org/10.1016/j.tre.2020.102067.
15. Bhutta, Muhammad Nasir Mumtaz, Amir A. Khwaja, Adnan Nadeem, Hafiz Farooq Ahmad, Muhammad Khurram Khan, Moataz A. Hanif, Houbing Song, Majed Alshamari, and Yue Cao. "A Survey on Blockchain Technology: Evolution, Architecture and Security." IEEE Access 9 (2021): 61048–73. https://doi.org/10.1109/access.2021.3072849.
16. Aggarwal, Shubhani, Rajat Chaudhary, Gagangeet Singh Aujla, Neeraj Kumar, Kim-Kwang Raymond Choo, and Albert Y. Zomaya. "Blockchain for Smart Communities: Applications, Challenges and Opportunities." Journal of Network and Computer Applications 144 (2019): 13–48. https://doi.org/10.1016/j.jnca.2019.06.018.
17. Zhang, Li, Yongping Xie, Yang Zheng, Wei Xue, Xianrong Zheng, and Xiaobo Xu. "The Challenges and Countermeasures of Blockchain in Finance and Economics." Systems Research and Behavioral Science 37, no. 4 (2020): 691–98. https://doi.org/10.1002/sres.2710.
18. Hjalmarsson, Friorik P., Gunnlaugur K. Hreioarsson, Mohammad Hamdaqa, and Gisli Hjalmtysson. "Blockchain-Based E-Voting System." 2018 IEEE 11th International Conference on Cloud Computing (CLOUD), July 2018. https://doi.org/10.1109/cloud.2018.00151.
19. Menachemi, Nir, and Collum. "Benefits and Drawbacks of Electronic Health Record Systems." Risk Management and Healthcare Policy, May 2011, 47. https://doi.org/10.2147/rmhp.s12985.
20. Iroju, Olaronke, Abimbola Soriyan, Ishaya Gambo, and Janet Olaleke. "Interoperability in healthcare: benefits, challenges and resolutions." International Journal of Innovation and Applied Studies 3, no. 1 (2013): 262-270.

21. Esmaeilzadeh, Pouyan, and Tala Mirzaei. “The Potential of Blockchain Technology for Health Information Exchange: Experimental Study From Patients’ Perspectives.” Journal of Medical Internet Research 21, no. 6 (2019): e14184. https://doi.org/10.2196/14184.
22. Sharma, Manu, and Sudhanshu Joshi. “Barriers to Blockchain Adoption in Health-Care Industry: An Indian Perspective.” Journal of Global Operations and Strategic Sourcing 14, no. 1 (2021): 134–69. https://doi.org/10.1108/jgoss-06-2020-0026.

Misinformation Detection in Text for COVID-19 Healthcare Data in South Africa

Seani Rananga[1][0000-0003-0984-7331], Nhlakanipho Ngwenya[2], Mahlatse Mbooi[3][0000-0002-8834-7513], Bassey Isong[4][0000-0002-3915-4627], Penelope Matloga[5], Vukosi Marivate[6][0000-0002-6731-6267]

[1]Data Science for Social Impact at University of Pretoria, Hatfield 0028, South Africa and Northwest University, Mahikeng 2970, South Africa
[1]seani.rananga@up.ac.za
[2]Data Science for Social Impact at University of Pretoria, Hatfield 0028, South Africa
[2] u18308342@tuks.co.za, [5]u22826476@tuks.co.za, [6]vukosi.marivate@cs.up.ac.za
[3]Council for Scientific and Industrial Research, Brummeria 0001, South Africa
[3]mratsoma@csir.co.za
[4]Northwest University, Mahikeng 2970, South Africa
[4] Bassey.Isong@nwu.ac.za

Abstract. The COVID-19 pandemic has witnessed an alarming expansion of misinformation, posing critical threats to public health. This research focuses on detecting misinformation within text data sourced from South African healthcare datasets during the COVID-19 crisis. The study aims to improve automated misinformation detection models tailored to the South African context by leveraging natural language processing (NLP) techniques and machine learning algorithms such as Logistic regression (LR). The investigation involves extensive analysis of datasets containing COVID-19 healthcare-related text, including the training and evaluation of machine learning models. In addition, NLP techniques, including LR, will be utilized to extract key features indicative of misinformation. The findings are poised to advance the development of effective tools and strategies to combat misinformation. Moreover, the study will adopt an inclusive approach by conducting analyses in both English and low-resource languages like isiZulu. This is critical to enhancing public health communication and strengthening defenses against the potential resurgence of COVID-19, thereby safeguarding public well-being.

Keywords: Natural Language Processing (NLP), Misinformation Detection, COronaVIrus Disease of 2019 (COVID-19), Logistic Regression (LR).

1 Introduction

The COVID-19 pandemic was not only a health crisis, but also fueled the widespread dissemination of misinformation through various communication channels, particularly social media. This misinformation often pertained to the treatment, vaccines, and prevention of COVID-19 [1]. In addition to reliable information on online platforms, there is an immense amount of misinformation circulating which can confuse, leading to unsafe health care decisions and distrust amongst society. To protect the public's health and ensure the accurate spread of information, it is essential to establish effective strategies for detecting and eliminating misinformation.

Misinformation is the unintentional spread of false information [1-2] while disinformation is the deliberate creation and sharing of false information with the intention to mislead or deceive [1-2]. Misinformation detection has grown to be a crucial component of global digital health care. This research topic has been covered on a global scale but not particularly in a context that is beneficial for the African population. This study focuses on the use of machine learning methods and natural language processing (NLP) [2] approaches to address misinformation in text concerning COVID-19 healthcare within the South African society. The dataset used in this research includes isiZulu and English languages, reflecting the linguistic diversity of the South African population. The research seeks to use systems that can automatically and accurately identify misinformation in a textual context.

The primary machine learning model employed in this study is the Logistic regression (LR) classifier [3], a deep learning model known for its effectiveness in sequence data analysis. The objective of this research is to explore and evaluate various methods and algorithms for detecting misinformation in COVID-19 healthcare-related text within online platforms such as X (formerly known as Twitter) in a South African context. By examining the existing literature, we aim to identify gaps in the current global knowledge, evaluate the effectiveness of different approaches, and propose innovative strategies for enhancing misinformation detection in the context of healthcare that effectively accommodates the South African population.

The proposed research will adopt a data-driven approach [4], utilizing large datasets of COVID-19 healthcare-related text from reputable sources, as well as datasets containing misinformation. We will evaluate models, with LR classifier as the primary model, that accurately differentiate between credible and non-credible information using innovative machine learning algorithms [5]. Transformer models such as BERT (Bidirectional Encoder Representations from Transformers) [6] and other deep learning models will be explored to detect misinformation in healthcare texts for future work. The outcomes of this research will support evidence-based decision-making, enhance public health communication, and mitigate the harmful effects of misinformation about the COVID-19 crisis. By developing reliable and efficient systems for misinformation detection, we can contribute to a more informed society and ensure the well-being of individuals and communities in South Africa.

The rest of the paper is structured as follows: Section 2 presents some of the related works on misinformation detection and Section 3 presents the methodology employed, while Section 4 and Section 5 present the findings and the conclusion, respectively.

2 Related Works

As the challenge of combating COVID-19-related misinformation and fake news intensifies, researchers have engaged in significant efforts to develop novel techniques and models for identifying and mitigating false information during the pandemic. This section reviews pertinent literature and contributions in this research domain. Detection of COVID-19 Misinformation: Hossain et al. [7] presented " COVIDLies," an approach to detecting COVID-19 misinformation on social media. Their work highlights the importance of real-time monitoring and analysis of user-generated content during health crises.

Machine Learning for Fake News Detection: Malla and Alphonse [5] employed ensemble pre-trained deep learning models for detecting informative tweets related to COVID-19. This approach showcases the application of machine learning in identifying and categorizing relevant content during the pandemic. Transformer-Based Models: Gundapu and Mamidi [6] proposed a transformer-based automatic COVID-19 fake news detection system. Their work explores the potential of transformer architectures in addressing the unique challenges posed by the infodemic.

NLP Techniques: Meystre et al. [4] emphasized the role of natural language processing in COVID-19 predictive analytics and data-driven patient advising. Their approach leverages NLP to provide data-driven insights amid the pandemic. Sentiment Analysis: Iwendi et al. [8] delved into sentiment analysis of COVID-19-related fake news. Their work showcases the importance of understanding the emotional context surrounding misinformation. Information Dissemination on social media: Shams et al. [2] introduced the" SEMiNExt" extension, a web search engine misinformation notifier, offering a machine learning-based approach to tackle misinformation during the COVID-19 pandemic. Their work underscores the value of real-time interventions.

This related work exemplifies the diverse range of approaches adopted by researchers in the domain of COVID-19 fake news detection. Machine learning, NLP, and sentiment analysis techniques have emerged as essential tools in identifying and combatting misinformation. These studies collectively contribute to the evolving landscape of solutions addressing the COVID-19 infodemic.

3 Methodology

The methods which are presently used to identify false information frequently rely on labor-intensive and error-prone manual fact-checking [9]. Although automated misinformation detection using NLP approaches and machine learning algorithms have demonstrated promising results, their usefulness in a South African COVID-19 healthcare context remains uncertain. Consequently, this research aims to evaluate a model for the effective identification of misinformation in the COVID-19 healthcare-related text dataset for South Africa. This research evaluates a model for the effective identification of misinformation in the COVID-19 healthcare-related text dataset for South Africa. To achieve this, we collected text data and pre-processed it, we then employed a logistic regression multinomial classifier, to train the model. The performance of the model was evaluated using metrics such as precision, recall, F1-score, and accuracy.

To expand our methodology, we have added the following subsections to evaluate more about the steps used in this study:

3.1 Dataset

An English dataset consisting of misinformation was adopted in this study and sourced from an article COVIDLies: Detecting COVID-19 misinformation on social media [10]. To uncover facts about the outcomes of the COVID-19 impact in South Africa, a factual website was used to source out facts about the impact of COVID-19 in South Africa from the platform Statista[1]. It is important to note that the low-resource language dataset for isiZulu does not exist, and hence, we did the translation process. We directly translated the English text into isiZulu, using the following translation approach: (1). The translation was conducted using Google Translate, (2) and the translations were cross-referenced using ChatGPT to ensure high-quality and accurate translations while maintaining a culturally sensitive and contextually appropriate translation. This translation approach diversifies the dataset by enabling the model to train on isiZulu text. Due to the amount of work it takes in translating the data, we only sampled a portion of English data to translate, leaving a development of a translation model of isiZulu as an open research gap, to ensure a larger amount of data to be translated at the same time.

3.2 Data loading and preprocessing

Once the Dataset is structured, cleaned, and organized, the data is then loaded from Comma-separated value (CSV) files, using Pandas to read the CSV file. The CSV file

[1] https://www.statista.com/statistics/1108127/coronavirus cases-in-south-africa-by-region/

has columns ‘Misinformation’ for statements and ‘Label’ for labels (1 for true, 0 for false). This CSV file combined both the English and isiZulu data for better accuracy of the model, due to lack of sufficient data for isiZulu. All words in the text are converted to lowercase.

3.3 Tokenization and Padding

The text data is then transformed into numerical form using tokenization, and sequence length consistency is maintained through padding. It is a type of feature engineering where the quality of text preprocessing and vocabulary selection determines how accurate the result is. Also, padding is used to guarantee that sequences are the same length for model training.

3.4 Data Splitting

This is a basic step in machine learning which involves dividing the dataset into three sets, the training, testing sets, and validation set. The quantity of data allotted to training and testing depends on the split ratio that is selected. This indicates that data dispersion was considered while evaluating the model. Therefore, the training data comprised 78% of the dataset, while the remaining 22% was equally split between the test and validation sets, with each receiving 11%.

The following sections will give the results of the study and give detailed explanations on the performances of the dataset defined in this methodology section.

3.5 Model architecture and training

We utilized a logistic regression model tailored for multi-class classification tasks. This model employs the 'lbfgs' solver, which efficiently manages smaller datasets and multicollinearity. It adopts a multinomial logistic regression approach, directly modeling multiple classes with the softmax function. During training, input features and target variables are prepared and passed to the model. The 'lbfgs' algorithm iteratively optimizes model parameters by maximizing the likelihood function, aiming to minimize the cross-entropy loss. This approach ensures robustness and efficiency in classifying text data, making it suitable for our study on isiZulu and English languages.

4 Results

This section focuses on evaluating the results, which is crucial for determining the model's performance and validity. Table 1 shows the Test set classification report, this gives a detailed breakdown of the classifier's performance for each class (in this case "False" and "True"). We used the precision, recall, f1-score, and support evaluation metrics. The precision for the "False" dataset is 0.88, which means 88% of instances predicted as False were False. For the "True" dataset, the precision was 0.75, this indicates that 75% of instances predicted as True were True. The recall(sensitivity) is the ratio of the true positive predictions to the total actual positives It indicates how many of the actual positive instances were correctly identified by the classifier. For "False": Recall performance was 0.70, this indicates that 70% of actual False instances were correctly identified. For "True" the Recall performance was 0.90, this indicates that 90% of actual True instances were correctly identified. F1-score is the harmonic mean of precision and recall. It provides a balance between precision and recall and is useful for evaluating the overall performance when there is an uneven class distribution. For "False": F1-score = 0.78 and for "True": F1-score = 0.82. Support indicates the number of actual occurrences of each class in the test set. For "False": Support = 10 and for "True": Support = 10.

Table 1. Test Set Classification Report

	Precision	**Recall**	**F1-score**	**Support**
False	0.88	0.70	0.78	10
True	0.75	0.90	0.82	10
Accuracy			0.80	20

Overall, the model demonstrates satisfactory performance with an 80% accuracy across 20 test instances. Notably, it excels at identifying true instances (90% recall) while maintaining reasonable precision (88%) for false instances. Further refinement could focus on improving precision for false instances without compromising recall.

Table 2 shows the validation set classification report, this consists of a validation set which is a subset of data used to evaluate how well a machine learning model performs on unseen data. The model shows high recall (1.00) for class 1.00, indicating it found all positive cases. However, the precision (0.47) indicates that less than half of the predicted positive cases were correct. The F1-score (0.64) balances these aspects, indicating moderate overall performance for this class.

With an overall accuracy of the model of 60%, the validation results in this table show that the model is performing poorly. It is only finding about half of the class 1.00 instances, and it is making a lot of mistakes in the other class.

Table 2. Validation Set Classification Report

	Precision	**Recall**	**F1-score**	**Support**
False	1.00	0.47	0.64	15
True	0.38	1.00	0.56	5
Accuracy			0.60	20

In summary, we presented the results of training the COVID-19 dataset for both English and isiZulu datasets combined. It was noted during this study that training data quality and quantity influenced the training data. Therefore, for future research, having more diverse and larger datasets will help improve the classifier's ability to generalize the performance. Most existing models do not cater to the natural language nuances, leading to performance trade-offs while working with low resource languages. To improve performance, future work should try a different machine learning algorithm and tune the hyperparameters of the model.

5 Conclusion

In this study, we used a logistic regression model to identify misinformation in text data from South Africa, focusing on isiZulu and English. The results show good overall performance, but challenges remain due to the lack of sufficient low-resource datasets. The dataset was divided into training, testing, and validation phases, with validation performance suffering from limited data availability. This research adds to the body of knowledge on machine learning methods for tackling COVID-19 misinformation in low-resource languages, highlighting the need for expanded capabilities in multilingual settings like South Africa. The main limitations include the small size of the dataset and limited model generalizability.

Looking ahead, we plan to extend the generalizability of these models to other languages and healthcare domains and to enhance model transparency by incorporating explainability techniques. These efforts aim to make machine learning a more effective tool for public health communication and misinformation management.

We also intend to refine our models with COVID-19-specific data to improve accuracy in identifying pandemic-related misinformation. Incorporating human expertise could further enhance detection capabilities. We are exploring the development of a real-time misinformation monitoring system and creating user-friendly apps to help verify COVID-19 information credibility. Such initiatives are essential for protecting public health and ensuring accurate information dissemination during ongoing global crises.

References

[1] S. Tasnim, M. M. Hossain, and H. Mazumder, “Impact of rumours and misinformation on COVID-19 in social media,” Journal of Preventive Medicine and Public Health, vol. 53, no. 3, pp. 171–174, 2020.

[2] A. B. Shams, E. Hoque Apu, A. Rahman, M. M. Sarker Raihan, N. Siddika, R. B. Preo, M. R. Hussein, S. Mostari, and R. Kabir, “Web search engine misinformation notifier extension (seminext): A machine learning based approach during covid-19 pandemic,” in Healthcare, vol. 9, p. 156, MDPI, 2021.

[3] J. Schmidhuber, S. Hochreiter, et al., “Long short-term memory,” Neural Comput, vol. 9, no. 8, pp. 1735–1780, 1997.

[4] S. M. Meystre, P. M. Heider, Y. Kim, M. Davis, J. Obeid, J. Madory, and A. V. Ale kseyenko, “Natural language processing enabling covid-19 predictive analytics to support data-driven patient advising and pooled testing,” Journal of the American Medical Informatics Association, vol. 29, no. 1, pp. 12–21, 2022.

[5] S. Malla and P. Alphonse, “Covid-19 outbreak: An ensemble pre-trained deep learning model for detecting informative tweets,” Applied Soft Computing, vol. 107, p. 107495, 2021.

[6] S. Gundapu and R. Mamidi, “Transformer-based automatic COVID-19 fake news detection system,” arXiv preprint arXiv:2101.00180, 2021.

[7] T. Hossain, R. L. Logan IV, A. Ugarte, Y. Matsubara, S. Young, and S. Singh, “Covid lies: Detecting covid-19 misinformation on social media,” in Proceedings of the 1st Workshop on NLP for COVID-19 (Part 2) at EMNLP 2020, 2020.

[8] C. Iwendi, S. Mohan, E. Ibeke, A. Ahmadian, T. Ciano, et al., “Covid-19 fake news sentiment analysis,” Computers and electrical engineering, vol. 101, p. 107967, 2022.

[9] F. Alam, S. Shaar, F. Dalvi, H. Sajjad, A. Nikolov, H. Mubarak, G. D. S. Martino, A. Abdelali, N. Durrani, K. Darwish, et al., “Fighting the covid-19 infodemic: modelling the perspective of journalists, fact-checkers, social media platforms, policymakers, and the society,” arXiv preprint arXiv:2005.00033, 2020.

[10] A. Glazkova, M. Glazkov, and T. Trifonov, “g2tmn at constraint@ aaai2021: exploiting ct-bert and ensembling learning for covid-19 fake news detection,” in Combating Online Hostile Posts in Regional Languages during Emergency Situation: First International Workshop, CONSTRAINT 2021, Collocated with AAAI 2021, Virtual Event, February 8, 2021, Revised Selected Papers 1, pp. 116–127, Springer, 2021.

9 Misinformation Detection in Text for COVID-19 Healthcare Data in South Africa

A Technology Based Model for Problem-Solving Skills Development in the Intermediate Phase

Roslyn Tait[1[0000-0002-5410-1881]], Dieter Vogts[2[0000-0002-2554-7518]] and Jean Greyling[3[0000-0002-6773-9200]]

[1, 2, 3]Department of Computing Sciences, Nelson Mandela University, Gqeberha, South Africa
s220316457@mandela.ac.za, dieter.vogts@mandela.ac.za,
jean.greyling@mandela.ac.za

Abstract. An educational model to improve problem-solving skills among learners is introduced. The literature review emphasises the importance of these skills, the impact of smart classrooms, and the influence of colour on learning. The model offers guidance for educators and learners to foster skills development in the intermediate phase through a structured workflow. A set of metrics to aid in the selection of tools to improve problem-solving skills is proposed. Based on the model, a proof-of-concept Internet of Things based system was successfully implemented and evaluated using a metric-based evaluation based on the proposed metrics. This evaluation highlights the effectiveness of the model in the implementation of a system in limited infrastructure and diverse educational settings.

Keywords: Problem-Solving, Internet of Things (IoT), RFID.

1 Introduction

In the twenty-first century, the development of children's skills across cognitive, interpersonal, intrapersonal, and technical domains has gained interest [1]. Within the cognitive domain, problem-solving skills, critical thinking, and computational thinking are considered fundamental in early childhood. Educators play an important role in nurturing these skills but often face challenges when attempting to incorporate these skills into their curriculum.

The concept of problem-solving encompasses various aspects and serves as a cognitive activity and a learning goal. Polya's problem-solving approach continues to be influential in the field of education [1, 2]. This process involves four steps: understanding the problem, devising a plan, carrying out the plan, and reflecting on the solution. Contrary to expectation, problem-solving is not meant to be frustrating; instead, it is meant to be a challenging task that promotes skill development and conceptual understanding [3].

Computational thinking has been said to promote problem-solving skills and equip learners with the ability to analyse and solve real-world problems effectively [4]. It provides a framework for problem-solving that can be used in various subjects to enhance a learner’s understanding and develop their problem-solving skills.

Computational thinking offers a systematic approach to formulating and solving complex problems. This mental process involves breaking down problems, abstracting them into computable forms, and leveraging computational tools to find solutions. The integration of computational thinking into education enhances critical thinking and problem-solving abilities, preparing individuals for the demands of the digital era.

Problem-solving is essential in the Intermediate phase, where learners are expected to develop a deeper understanding of life skills [5]. The Intermediate phase includes grades 4 to 6, which encompasses learners aged 9 to 12.

In 2019, the Trends in International Mathematics and Science Study (TIMSS) assessed the mathematics proficiency of fourth-grade learners in 64 countries [6]. TIMSS uses four points along its scale as International Benchmarks to interpret results: Advanced (625), High (550), Intermediate (475), and Low (400). Most countries had less than 10 percent of fourth grade learners performing at the Advanced level. The international median percentage of learners reaching each benchmark were 7 percent for Advanced, 34 percent for High, 71 percent for Intermediate, and 92 percent for Low. Within this study, South Africa (SA) ranked third lowest in mathematics achievement, scoring 374, while the highest-performing country, Singapore, scored 625. The TIMSS achievement scale sets a centerpoint of 500 as the mean. SA's score falls significantly below this and there is a notable gap when compared to the top-performing country.

Additionally, a study by the Institution of Engineering and Technology (IET) in 2019 found a 10-14% decline in interest in most STEM subjects among 9–12-year-olds compared to 2015 [7]. This decline in global interest highlights the need to focus on the Intermediate phase. This reveals a lack of crucial skills and interest needed for subjects relying on these skills at this age. Improving problem-solving skills and fostering interest in these subjects could enhance the overall performance of learners within this age group.

This paper proposes a technology-based model for problem-solving skills development. It reviews the existing literature on smart classrooms, tools for problem-solving skills development and the impact of colour on learning (Section 2). The design of the proposed model is presented in Section 3 and a proof-of-concept implementation of the model is discussed in Section 4. Section 5 reports on the evaluation of the implemented system, against specific predefined metrics (Section 2.2). Finally, Section 6 summarises the findings of this paper and provides recommendations for further research.

2 Literature Review

By reviewing relevant literature, this section aims to discuss the concept of smart classrooms and their use of technology to improve essential 21st-century skills such as problem-solving. It also discusses the selection of technology tools for the improvement of problem-solving skills and explores the influence of colour on learning.

2.1 Smart Classrooms

The Internet of Things (IoT) has transformed several industries, including education [8]. With the increasing connectivity between objects, environments, and people, IoT promises to offer a fully connected and "smart" world [9]. This has also led to the implementation of smart classrooms.

A smart classroom, also known as an intelligent classroom, future classroom, or technology-enhanced classroom, represents a dynamic educational environment using advanced technologies to optimise teaching and learning processes [10]. While there is currently not a universal definition, various perspectives highlight its key features and functionalities. In general, a smart classroom is an innovative learning space that integrates information and communication technology (ICT) to enhance the educational experience. It is an environment where traditional teaching methods are used in addition to advanced technology to create an engaging and interactive learning environment. Integration in education includes the use of smart devices, sensors, and wearable technology in classrooms. These devices can collect data and provide real-time feedback, which allows educators to personalise instruction and track learner progress effectively.

Smart classrooms prioritise learner-centred teaching and aim to accommodate diverse learning styles and abilities, foster lifelong learning, and support ongoing development. These environments represent the evolution of traditional educational settings into technology-enabled spaces that improve essential 21st-century skills such as communication, critical thinking, problem-solving, creativity, and collaboration. Overall, smart classrooms help learners develop essential skills necessary for success in the modern world and introduce learners to the use of ICT.

2.2 Tools for Problem-Solving Skills Development

In the digital age, educators have a variety of technological tools at their disposal to enhance classroom instruction, facilitate learner engagement and develop learners' skills. With educators incorporating digital learning tools into their daily instruction, selecting the right tools is important. However, the large number of available tools creates a challenge for educators in selecting the most appropriate tools, understanding their value in education, and knowing how they can integrate the tools into their classrooms [11, 12].

TPACK or Technological Pedagogical Content Knowledge is an understanding that develops from interactions between content, pedagogy, and technology knowledge. SAMR (substitution, augmentation, modification, and redefinition) is a framework of ordered strategies for technology integration that seeks to help educators create an engaging and transformative learning experience. TPACK and SAMR are useful integration frameworks that provide educators with a structured approach to effectively incorporating technology into instruction [12]. However, the frameworks focus more on helping educators think about how to apply technology in the classroom rather than how to select technology tools. In order to select the most appropriate tools for their classrooms, educators need to compare a variety of available tools.

Currently, various online resources are available that provide a list of available technology tools that can be used by educators [11, 13–15]. These resources provide educators with a list of tools, their descriptions, and a link to the given tool. Additionally, various registries of tools and their descriptions exist, such as SEEK-AT-WD. SEEK-AT-WD leverages a social-semantic approach to curate tool descriptions, enabling educators to access up-to-date information and share their experiences with different tools [16]. While useful for finding educational technology tools, these registries do not assist educators in tool selection. There are various resources to assist educators in selecting the most valuable technology tools for their classrooms. These resources include guides, checklists, tips, considerations, or questions to help educators make informed decisions about which tools to select [12, 17–19]. These emphasise factors such as alignment with learning objectives, usability, accessibility, and long-term benefits. The resources provide an outline for educators to use to make informed decisions about which tools to select. However, with such a wide variety of sources, it can be difficult and time-consuming to evaluate tools based on all suggested criteria. Additionally, these evaluations do not specifically target problem-solving skills.

Rich et al. [20] propose that affordability and accessibility play a significant role in selecting tools used to teach computational thinking, stating that “As both teachers and researchers make choices regarding tools, they must weigh the current and future costs of the tools as well as the affordances provided for a diverse body of students”.

As highlighted, limited literature exists on criteria to evaluate tools for problem-solving skills development. In addition to literature, input was received during an interview with Bronwen Jonson in November 2023. She is a primary school teacher from Summerwood Primary School. The following metrics are proposed:

- **Affordability:** In educational settings, budget constraints may limit educators' access to certain tools. A cost-effective tool may increase accessibility and adoption among educators and learners [20].
- **Accessibility/Availability:** Ensuring that tools are accessible to all learners, regardless of their socio-economic background or physical abilities, is important. Tools with low barriers to access can promote inclusivity and equitable learning opportunities for all learners. Whether the tool requires internet access will have an effect on the accessibility of the tool [20].
- **Ease of use:** User-friendly interfaces and intuitive design can enhance learner engagement and reduce the learning curve for educators [Interview].
- **Compatibility with existing infrastructure:** This is essential for integration of the new tool into the classroom environment. Tools that require minimal additional equipment or technical support may be more likely to be adopted and sustained over time in certain learning environments [Interview].
- **Group size:** Knowing the number of learners the tool can accommodate, is essential for determining its effectiveness in each scenario [Interview].
- **Target age** [Interview].

These metrics will be used in Section 5 when evaluating a Proof-of-Concept system.

2.3 Impact of Colour on Learning

The effective use of colour in learning environments is a valuable design element that adds definition to a space and improves the overall aesthetics of a classroom [21, 22]. In addition to this, the use of colour in learning materials enhances learners' reception of information and makes them pay more attention to and recall information better [21, 23]. In learners, colour also triggers physical, emotional, and cognitive effects that can have a positive or negative influence on learners' mood, feelings, attention, productivity, communication, performance, and achievement [24].

Colour for colour's sake is not constructive, and it is important to carefully consider the use of colours to create an environment that is not overstimulating or over-tranquilising for learners [21, 22, 24]. Large quantities of colours should be avoided, and the application of colours should be balanced. Doing so will help create an environment conducive to learning and specific focal points for activities within the classroom.

Different colours have connotations and evoke various reactions from learners [23, 24]. Bright colours stimulate and motivate, while dark colours may evoke negative feelings. Light colours like yellow and light blue energize learners and demand attention. Green and blue promote relaxation and calmness, ideal for overactive learners. Red and orange should be used sparingly when drawing learners' attention to specific information, as they have been known to make some learners anxious.

Effective use of colour in classrooms enhances learning by engaging and motivating learners. Educators can adapt their environments and materials based on the effects of different colours. Balancing these colours creates an environment where learners are captivated and find it easier to absorb information. Overall, this approach creates a blend of visual stimulation and cognitive support to enhance learning experiences.

3 Design

This research paper proposes an educational model to enhance learners' skills through hands-on learning. This model relies on both educator and learner interaction to complete a task. Within this model, learners are encouraged to design an IoT-based system. This system is based on the concept of Blocks, which are interconnected components. In the context of this paper, the task is not the primary focus, as it can be modified to suit multiple scenarios and academic concepts, but it is important for implementation purposes.

3.1 Task

A task within the context of this paper is an activity or assignment developed to impart knowledge onto learners or develop certain skills. Problem-solving and critical thinking are considered important skills that this task should improve. While the nature of the task itself is not the primary focus of this paper, it is necessary to provide a task to demonstrate the application of the model. Tasks can take various forms to actively involve learners in the learning process. They aim to encourage learners to apply knowledge, analyse information, and develop creative solutions to problems. The selection and design of tasks play a pivotal role in shaping the learning experience and outcomes. Tasks should be aligned with the learning objectives and can be scaffolded to accommodate learners' diverse needs or grade levels. Educators can enhance learners' motivation, engagement, and knowledge retention by incorporating authentic and meaningful tasks into the curriculum.

Tasks could range from playing educational games like Mastermind to conducting data collection for generating reports or designing devices for specific activities. Some examples of tasks could include instructing learners to do the following:

1. Mastermind – play the game mastermind and report the results.
2. Weather report – create weather reports in different environments. Learners should use relevant sensors to generate a report, which can be analysed and presented.
3. Light spectrum analysis – explore the properties of light and colour. Learners should use sensors to measure and record readings from different light sources or objects. This data can then be used to plot the intensity of light at different wavelengths (red, green, blue) helping to visualize the spectrum.
4. Distance measurement – build and use a measuring device to explore the distances to objects in different environments. This can be used as an introduction to distance measurement, to help learners estimate distances, or to perform conversions between different distance units. Learners should record, analyse, and present their results.

While the specifics of tasks may vary based on educational context, they should all promote learning and problem-solving skills development among learners.

3.2 Blocks

The IoT-based system proposed in this paper revolves around the concept of Blocks, which are interconnected components used to create a system to complete a task. A Block refers to an independent physical unit with at least one function, such as collecting data, processing data, or producing outputs. Blocks are able to perform multiple functions.

Each Block is equipped with hardware components and software to perform its required functions. The hardware of the Block is a microcontroller unit (MCU) connected to sensor, input, or output devices. The specific hardware configuration of a Block is dictated by its intended function. By linking multiple Blocks together, users can create custom systems within the smart environment. This connectivity is crucial for facilitating communication among Blocks, enabling the collection and exchange of data.

The software embedded within each Block is loaded onto the MCU and facilitates the Blocks processing and inter-Block communication. Wireless connectivity is employed to ensure a robust, easily configurable system. Moreover, the absence of physical wires simplifies the Block design, making it more accessible to younger users.

To enable inter-Block communication, it is important to select an IoT communication protocol. IoT communication protocols are necessary to ensure efficient, secure, and dependable data exchange between devices [25]. Adhering to a common protocol within a system enables a seamless integration of devices. Each IoT communication protocol has unique capabilities, functionalities, and distinct characteristics, such as transmission range, resource consumption, and power usage. Wireless IoT communication protocols specifically offer several advantages over wired protocols, including scalability, interoperability, and reliability. Wireless IoT communication protocols can be categorised into short-range and long-range communication protocols. Short-range protocols, such as Bluetooth and Zigbee, operate within a limited range, lowering connectivity costs and power consumption. Long-range protocols, such as LoRaWAN, are designed to cover larger distances, often reducing throughput to conserve power for long-distance transmissions.

Given the system's constraint of Blocks within the same environment, only short-range protocols were considered. Among these was Bluetooth Low Energy (BLE), an optimised version of Bluetooth, which is a wireless technology that enables high-speed data exchange over short distances in small amounts. While BLE has better energy efficiency than Bluetooth, BLE still has some drawbacks as its data transfer speed is slower, it still uses the crowded 2.4 GHz frequency, and it is not always suitable for large files. Despite these drawbacks, there are several advantages to BLE, namely low latency, lower implementation costs due to hardware simplicity, default data encryption, and the fact that it does not require an expensive custom gateway to control connected devices. Due to these advantages and limited drawbacks, BLE was selected as the preferred protocol for this system.

Block colours were selected to represent the functionality of each Block, namely input, processing, or output. In cases where Blocks contained more than one function, the predominant function’s colour was chosen. Following the colour theory discussed in Section 2.3, input Blocks were designated as blue, processing Blocks as green, and output Blocks as yellow. Additionally, power Blocks were red to draw attention to their importance and their scarcity ensures that the colour is not overwhelming.

The tasks described in Section 3.1. can be implemented using various Blocks. For example, an input Block in the form of a joystick providing up, down, left, right, and click inputs can be used for task 1. Tasks 2 and 3 could use an input Block made up of various environmental sensors, including temperature, light, water level, and colour sensors. For task 4, an input Block in the form of a distance sensor measuring in millimetres, centimetres, meters, and kilometres could be used. Additionally, a processing Block known as a sensor report generator could be used to process data for readable reports for Tasks 2, 3, and 4. Output Blocks could include various displays such as a standard 16x2 LCD screen and a low-power 0.96-inch OLED screen, both would be applicable to all tasks. Furthermore, power Blocks would be required for all tasks.

3.3 Proposed Model

The proposed model aims to guide educators and learners on the process to promote skills development in the intermediate phase. In **Fig. 1**, the process to be followed by educators is shown in blue, and the subprocess to be followed by learners is shown in yellow. The learner process is seen as an educator subprocess, as steps need to be completed by the educator before the learners are able to start their process.

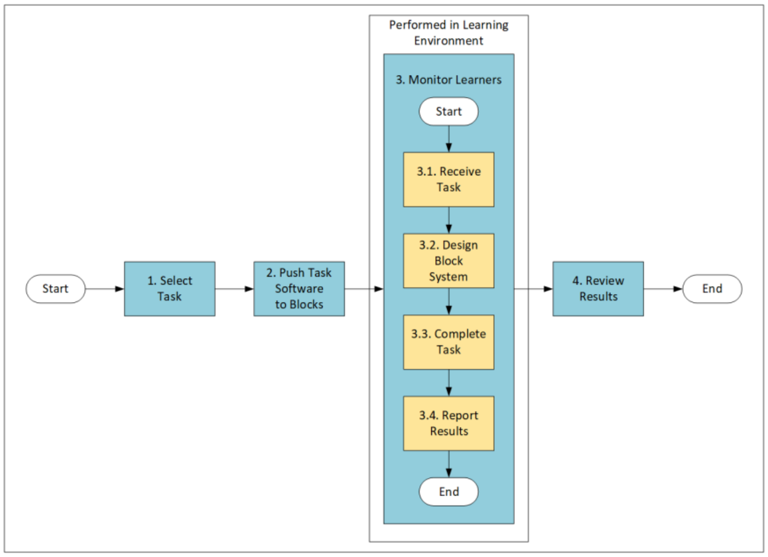


Fig. 1. Proposed Educational Model

The educator workflow is as follows:

1. Select Task: The educator starts the workflow by selecting a task from a predetermined list of tasks. Educators will be provided with a detailed description of the task and a list of Blocks that may be used to complete the task. The educators will also be given the necessary software for the Blocks for the specific task. It is necessary to provide educators with the software as Blocks may be used in multiple tasks and will therefore require task-dependent software.
2. Push Task Software to Blocks: Once the educator has selected a task and has the necessary 'Task Software', they will need to load the software onto the Blocks. This is known as 'pushing' the software. To push the software onto the Blocks, educators should be able to load the code onto the Blocks via a USB cable connecting the Block to their computer or using an over-the-air (OTA) software update to wirelessly upload the new 'Task Software' onto the Blocks.
3. Monitor Learners: This step is performed within the learning environment once the software has been loaded on the Blocks. Educators are able to start the learner workflow by providing learners with the task to complete and a selection of Blocks. Educators should then monitor learners and provide guidance, if necessary, until learners have completed the task. It may be necessary for educators to stop the learner workflow prematurely due to time constraints, but this can be done at the educator's discretion.
4. Review Results: After learners have completed their workflow and provided their results to the educator, the educator is able to review and track the progress of learners. This will give the educator insights into each learner's performance.

The learner workflow is as follows:

3.1. Receive Task: The learner process begins when they receive a task from their educator. These tasks can vary, covering subjects such as programming exercises, scientific experiments, or general critical thinking and problem-solving challenges. The method by which tasks are distributed to learners can also vary. Educators can verbally communicate the task to learners or provide a visual aid such as a printout with the task description or an instructional video.

3.2. Design Block System: Upon receiving the task, the next step is to design an IoT-based system to perform the given task. This involves selecting and configuring the appropriate Blocks to create a system capable of completing the task. This step promotes problem-solving and encourages learners to think about the task in more detail before attempting to complete it.

3.3. Complete Task: With the system in place, learners may attempt to complete the task using their designed system. During this step, educators may also encourage learners to communicate with each other to foster teamwork and collaboration.

3.4. Report Results: Finally, when learners have completed the task, they should report their results. This should be done by allowing learners to verbally report their findings and whether or not they were able to complete the task. Educators should record each learner's results to monitor progress.

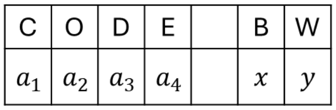
4 Implementation

For the purposes of this paper, implementation focuses on the technical implementation of the Block components. The task selected for this paper was Mastermind. All implementation details were based around this task to serve as a proof of concept for the learner workflow in the proposed model, given in Section 3.3.

4.1 Mastermind

The game "Mastermind" was selected as the task due to its ability to develop the problem-solving, critical thinking and scientific reasoning skills of players [26, 27]. The game is played when the device acting as the code maker creates a secret code, which the codebreaker attempts to determine in as few guesses as possible. Each guess consists of entering an ordered sequence of 4 symbols into the code maker device. These symbols are selected from a set of six possible options. In this case, the options are Heart, Bell, Smile, Person, Arrow, and Lock.

The code maker will then provide feedback on the guess by placing black and white pegs to indicate the accuracy of the guess. A black peg indicates a correct symbol in the correct position, while a white peg indicates a correct symbol in the wrong position. In this case, a peg is placed by updating a counter for the respective peg.



C	O	D	E		B	W
a_1	a_2	a_3	a_4		x	y

Fig. 2. Mastermind guess format.

Fig. 2 illustrates the format of a Mastermind guess with feedback. The guess is indicated by the sequence $\{a_1, a_2, a_3, a_4\}$ and the counters for black and white pegs are indicated by x and y, respectively. This format allows for all the necessary information to be displayed to the user in a minimal amount of space.

4.2 Implementation of Mastermind

To play the game of Mastermind, two Blocks are required, in addition to the generic power Block. The first is an input Block that can provide four directions (up, down, left, right) and a button click to indicate the submission of a guess. The up and down directions are necessary to scroll through the symbol options available for the current position, whereas the left and right directions are necessary to move between the symbols in the sequence. The second is an output Block responsible for displaying the necessary user interface (UI) shown in **Fig. 2** on a screen. Due to the simple nature of the game, minimal processing is required for execution. This allows the processing to be done within the output Block, thereby simplifying the system.

Fig. 3 illustrates the implementation of the Blocks needed to complete the Mastermind task. In this image, it can be seen that three different coloured Blocks have been created. The power Block is given in red and has a white battery symbol on top to indicate its function. The input Block is shown in blue and includes a joystick module which takes input from the user in the form of a direction or click. Finally, the output and processing Block is given in yellow and provides the graphical interface and processing necessary to complete Mastermind. Each Block has a Lego casing comprised of multiple Lego blocks. This casing is needed to show the Block colour and protect the hardware components of the Block. Lego was selected as a cheaper alternative to a 3D-printed case. Additionally, Lego is more robust and allows for customisation through the addition of other Lego blocks or creations, such as mini figures.

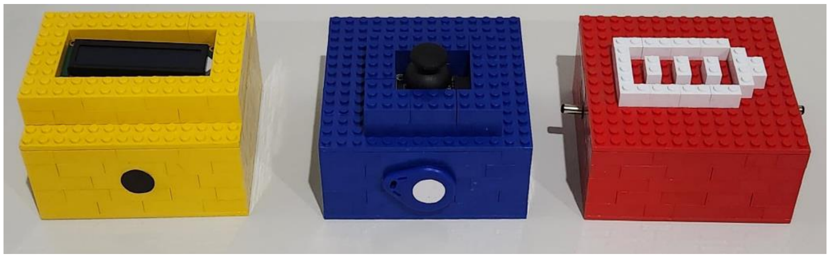


Fig. 3. Implementation of Blocks.

The power Block casing, as seen in **Fig. 3**, serves as a protective housing for multiple 9V batteries. It features output terminals protruding from the casing, enabling the Block to connect to other Blocks. This Block functions as an external power supply for other Blocks, offering an alternative to individual 9V batteries, which might be daunting for younger learners. Batteries were chosen as the power source due to their replaceability, accessibility, ease of use, and ability to sustain the MCU for extended periods due to the low power consumption of BLE.

Both the input and output Blocks make use of an MCU to facilitate data collection, exchange, and processing. The MCU selected for the Blocks was the ESP32 DEVKIT V1 board, a low-cost and low-power consumption microcontroller with integrated Wi-Fi, Bluetooth and BLE. It was chosen due to its high processing performance, reduced cost, and adaptability.

In Section 3.2, BLE was chosen as the preferred communication protocol. BLE was implemented due to the advantages previously discussed, and it does not rely on internet access or other components for communication. In order to make use of BLE to connect the Blocks, they needed to determine which device to connect to.

In this implementation, the input Block functions as a BLE server that advertises a unique service UUID. This service UUID was generated to be a unique identifier that indicates that the Block is an input Block accepting directions (up, down, left, or right) or a click as input from the user and broadcasting it to any connected Blocks. When started, this server will create a service with the UUID and advertise it to all BLE-enabled devices. Once connected to another Block, the input Block will be able to broadcast the user input to the connected device.

The output Block within this implementation functions as a BLE client looking to connect to a BLE server capable of providing user input in a specific format, namely any input Block broadcasting directions (up, down, left, right) or a click. This Block is able to connect to any service providing the necessary input, meaning it contains a list of service UUIDs that it is able to connect to. Therefore, if another input Block is advertising alongside the joystick input Block in the same environment, this client is able to connect to either device as long as their UUIDs are contained in the client's list of valid service UUIDs.

In this implementation, determining how the output Block would connect to an input Block posed a challenge. In order to connect the Blocks, the UUID of the Block would need to be pre-configured, or the user would be required to select the input Block they wanted to use. Typically, this is done by allowing the user to select a device to connect to via a UI. While this would be possible using the hardware of the implemented output Block, it may not be possible on other output Blocks such as a Block with an LED matrix. This prompted the development of an alternative way to facilitate the selection of an input Block.

The alternative connection strategy uses RFID tags and readers. In this strategy, Blocks are equipped with a tag, a reader or both, depending on their function. Blocks with readers can establish connections to Blocks with tags by simply 'tapping' them together. The process of 'tapping' is when the RFID tag, visible to the user, is read by the RFID reader indicated on the Block by a sticker, as seen in **Fig. 3**. Tapping the Blocks together signals to the Block with the reader, acting as the client, that the Block with the tag, acting as the server, wants to set up a connection. To determine which server wants to connect when multiple devices are advertising in the network, the tag is encoded with the unique service UUID of the server.

In this implementation, the input Block has an RFID tag attached to its casing, encoded with the Blocks unique service UUID. Additionally, the output Block has an

RFID reader within its casing that can read the tag. This output Block will wait for an input Block to be connected before allowing the Mastermind task to start. When an input Block with a valid service UUID is tapped against the output Block, a connection will be established between the Blocks, and the Mastermind task will start automatically. The output Block will generate a secret code, display the Mastermind UI to the user, and wait to receive inputs before updating the UI.

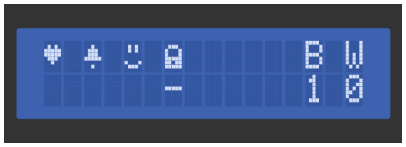


Fig. 4. Implementation of Mastermind UI.

Fig. 4 illustrates the implemented Mastermind UI. The four symbols to the left of the top row indicate the user's current guess, while the “-” symbol on the bottom row indicates the symbol in the code that the user is currently selecting. The numbers below the “B” and “W” represent the number of black and white pegs, respectively. This UI will be updated based on the inputs received from the input Block. When the user inputs a click, the current guess will be evaluated, and the number of black and white pegs will be updated. Once the correct code is provided, the task is considered complete, allowing the learner to report their results and end their workflow.

5 Evaluation

A metric-based evaluation was conducted to assess the implemented system, as discussed in Section 4. This evaluation technique was chosen as usability testing has not yet been conducted. The evaluation was based on the metrics discussed in Section 2. The review aimed to analyse the system's performance in terms of affordability, accessibility/availability, ease of use, and compatibility with existing infrastructure, as well as considering the number of learners the tool can accommodate, and the target age group.

5.1 Affordability

In order to evaluate the affordability of the system, a detailed cost analysis was performed. The total cost of the system amounts to R 1 083,99, with each of the three Block components incurring individual costs as follows: Power Block (R 255,70), Input Block (R 369,12), and Output Block (R 459,17). The full breakdown of the costs can be seen in **Table 1**, **Table 2**, and **Table 3**.

Table 1. Cost of Power Block

Component	Cost (ZAR)
2 x DC Power Jack Connectors (Male)	23,80
2 x 9V Batteries	159,90
Lego	63,98
Etc.	10,00

Table 2. Cost of Input Block

Component	Cost (ZAR)
ESP32 DEVKIT V1 Board	199,95
Joystick Module	59,95
RFID Tag 13.56MHZ 1KB	9,95
2 x Mini Breadboards	11,30
Tactile Switch (12x12mm)	7,92
DC Power Jack Connector (Female)	8,05
Lego	63,98
Etc.	10,00

Table 3. Cost of Output Block

Component	Cost (ZAR)
ESP32 DEVKIT V1 Board	199,95
LCD I2C Display	89,95
RFID-RC522 Module	70,00
2 x Mini Breadboards	11,30
Tactile Switch (12x12mm)	7,92
DC Power Jack Connector (Female)	8,05
Lego	63,98
Etc.	10,00

The provided breakdown offers a transparent insight into the financial aspect of the system's development, allowing for informed decision-making and resource allocation strategies within budget constraints.

5.2 Accessibility/Availability

The system uses solid-coloured Blocks in engaging colours (yellow, blue, red) to prevent overstimulation. This design choice positively impacts learning experiences by enhancing engagement and comprehension. All electronic components utilised in the system are widely available, ensuring easy procurement and accessibility. Furthermore, the system functions without the need for internet access, enhancing accessibility in environments with limited or no internet connectivity. It can accommodate one or multiple learners, facilitating collaborative learning experiences. Blocks can be reused in new tasks, making the system more adaptable and accessible in various learning environments. The system's design promotes broader access for learners from various backgrounds, promoting overall accessibility and availability.

5.3 Ease of Use

The system features a user-friendly and intuitive interface, especially for learners familiar with Mastermind. Educators can easily integrate the explanation of the Mastermind interface into their workflow, which in turn can facilitate the seamless adoption by learners. The use of a joystick for input allows for quick interactions [28]. Its flexibility and user-friendly design also ensure it is comfortable and easy to use. This helps mitigate difficulties learners may face with traditional button-based interfaces.

5.4 Compatibility with Existing Infrastructure

The system requires a computer or mobile device for uploading code. This additional equipment is necessary only if new software is needed for the Blocks. While this is a limitation of the system, if it is implemented in a smart classroom, it can be assumed that a computer or mobile device is available. In cases where this isn't available, the Blocks can be relocated to a place where such equipment is accessible, reprogrammed, and then returned. With no reliance on network infrastructure or specialised equipment, the system seamlessly integrates into learning environments, particularly smart ones, ensuring compatibility and straightforward adoption.

5.5 Additional Metrics

The system's target age group is learners in the intermediate phase, namely 9 to 12-year-old learners. This aligns with learners' cognitive and developmental needs within this age range. The system enhances its relevance and effectiveness in educational settings by catering to this age group. The system is designed to accommodate one or more learners. When used by multiple learners, it fosters collaboration, communication, and problem-solving skills development.

6 Conclusion

An educational model designed to enhance problem-solving skills among learners is introduced. The paper highlights the importance of developing these skills, the impact of smart classrooms, the selection of tools for improvement of problem-solving skills and explores the influence of colour on learning. A list of metrics to assist in the selection of tools, is proposed. The proposed model is built upon these insights, offering guidance for educators and learners to foster skills development in the intermediate phase through a structured workflow.

As part of this model, the learner workflow focusing on system design and task completion was successfully implemented with a proof-of-concept system to complete the Mastermind task. This implementation was evaluated using a metric-based evaluation which highlighted the system's effectiveness across key metrics, including affordability, accessibility, ease of use, and compatibility. The system incorporates inclusive design features to improve learning experiences and foster the development of critical skills, including problem-solving. The system is able to function with limited existing infrastructure and without internet or electricity, due to the use of batteries. This is valuable in learning environments with limited resources, allowing it to be used in diverse educational settings. However, batteries have a limited operational lifespan and will require replacement or recharging once depleted. Additionally, the system requires a computer or mobile device for uploading code when new software is needed for the Blocks. While these are limitations, they have been considered and the current implementation remains the most suitable approach. The findings of this paper indicate that the implemented system promotes equitable and engaging learning opportunities while developing the problem-solving skills of intermediate-phase learners.

Future research should explore the implementation and evaluation of additional tasks and other aspects of the proposed model, such as the educator workflow. This paper offers a foundation for future research and provides opportunities for further exploration in subsequent research.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Granone, F., Kirsti Lie Reikerås, E., Pollarolo, E., Kamola, M.: Critical Thinking, Problem-Solving and Computational Thinking: Related but Distinct? An Analysis of Similarities and Differences Based on an Example of a Play Situation in an Early Childhood Education Setting. In: Teacher Training and Practice, pp. 77-94. IntechOpen, London (2023). doi: 10.5772/intechopen.110795
2. Pólya, G.: How to Solve It: A New Aspect of Mathematical Method. 2nd edn. Princeton University Press, Princeton, New Jersey (1971)
3. Thornton, S.: Children Solving Problems. Harvard University Press, Cambridge, MA (2009)
4. Main, P.: Computational Thinking, https://www.structural-learning.com/post/computational-thinking, last accessed 2024/03/08
5. CAPS 123: Teaching Essential Life Skills in the Intermediate Phase: Strategies and Examples, https://caps123.co.za/teaching-essential-life-skills-in-the-intermediate-phase-strategies-and-examples/, last accessed 2024/03/08
6. Mullis, I.V.S., Martin, M.O., Foy, P., Kelly, D.L., Fishbein, B.: TIMSS 2019 International Results in Mathematics and Science. TIMSS & PIRLS International Study Center, Boston College (2020)
7. The Institution of Engineering and Technology: Inspiring the next generation of engineers. (2019)
8. Tripathi, G., Ahad, M.A.: IoT in Education: An Integration of Educator Community to Promote Holistic Teaching and Learning. In: Nayak, J., Abraham, A., Krishna, B., Chandra Sekhar, G., Das, A. (eds.) Soft Computing in Data Analytics, Advances in Intelligent Systems and Computing, vol. 758, pp. 675-683. Springer, Singapore (2019). doi: 10.1007/978-981-13-0514-6_64
9. Al-Emran, M., Malik, S.I., Al-Kabi, M.N.: A Survey of Internet of Things (IoT) in Education: Opportunities and Challenges. In: Hassanien, A.E., Bhatnagar, R., Khalifa, N.E.M., Taha, M.H.N. (eds.) Toward Social Internet of Things (SIoT): Enabling Technologies, Architectures and Applications, Studies in Computational Intelligence, vol. 846, pp. 197-209. Springer, Cham (2020). doi: 10.1007/978-3-030-24513-9_12

10. Zhang, Y., Li, X., Zhu, L., Dong, X., Hao, Q.: What Is a Smart Classroom? A Literature Review. In: Yu, S., Niemi, H., and Mason, J. (eds.) Shaping Future Schools with Digital Technology, Perspectives on Rethinking and Reforming Education, pp. 25-40. Springer, Singapore (2019). doi: 10.1007/978-981-13-9439-3_2
11. Sphero: 25 Technology Tools for The Classroom, https://sphero.com/blogs/news/technology-tools-for-the-classroom, last accessed 2024/03/09
12. Moore, R.L.: Educational Technology Tools. In: David, M.E. and Amey, M.J. (eds.) The SAGE Encyclopedia of Higher Education, vol. 1, pp. 425-426. SAGE Publications, Inc., Thousand Oaks, California (2020)
13. Rozario, R.N.: 30+ Best Educational Technology Tools for 2024, https://wpmanagen-inja.com/best-educational-technology-tools-teachers-students/, last accessed 2024/03/09
14. Wahl, H.: 13 Free Technology Tools for Elementary Teachers, https://www.koda-ble.com/learn/free-technology-tools-for-elementary-teachers, last accessed 2024/03/18
15. Myers, H.: 27 Tech Tools Teachers Can Use to Inspire Classroom Creativity, https://ozo-bot.com/27-tech-tools-teachers-can-use-to-inspire-classroom-creativity/, last accessed 2024/03/18
16. Ruiz-Calleja, A., Vega-gorgojo, G., Asensio-Pérez, J., Gómez-Sánchez, E., Bote-Lorenzo, M., Alario-Hoyos, C.: SEEK-AT-WD: A social-semantic infrastructure to sustain educational ICT tool descriptions in the Web of Data. Journal of Educational Technology & Society **17**(2), 321-332 (2014)
17. eSpark: Guide to Evaluating Technology Tools for Teachers, https://www.esparklearn-ing.com/blog/evaluate-technology-tools-for-teachers/, last accessed 2024/03/09
18. Petrick, D.: 6 Tips for Choosing Educational Technology, https://www.wiley.com/en-us/net-work/education/instructors/teaching-strategies/6-tips-for-choosing-educational-technology, last accessed 2024/03/09
19. Renard, L.: Choosing the best classroom technology - 5 things teachers should think about, https://www.bookwidgets.com/blog/2020/02/choosing-the-best-classroom-technology-5-things-teachers-should-think-about, last accessed 2024/03/18
20. Rich, P.J., Bartholomew, S., Daniel, D., Dinsmoor, K., Nielsen, M., Reynolds, C., Swanson, M., Winward, E., Yauney, J.: Trends in tools used to teach computational thinking through elementary coding. Journal of Research on Technology in Education **56**(3), 269-290 (2022) . doi: 10.1080/15391523.2022.2121345
21. Datta, M.S.: The Impact of Colour. Humanities Graduate Research. (2008)
22. Read, M.: Designing with Color in the Early Childhood Education Classroom: A Theoretical Perspective. Creative Education **10**, 1070-1079 (2019). doi: 10.4236/ce.2019.106080
23. Taylor, M.: Color Psychology: How to Best Use 6 Colors in Learning, https://imagination-soup.net/color-psychology-how-to-use-color-in-learning-colorize/, last accessed 2024/03/08
24. Amarin, N., Al-Saleh, A.: The Effect of Color Use in Designing Instructional Aids on Learners' Academic Performance. Journal of E-Learning and Knowledge Society **16**(2), 42-50 (2020) . doi: 10.20368/1971-8829/1135246
25. Expanice: Expert Tips for Choosing an IoT Protocol for Your Project, https://expanice.com/ar-ticle/iot-communication-protocols-comparison, last accessed 2024/03/13
26. Strom, A.R., Barolo, S.: Using the Game of Mastermind to Teach, Practice, and Discuss Scientific Reasoning Skills. PLoS Biology **9**(1), 1-3 (2011). doi: 10.1371/journal.pbio.1000578
27. Bach, M.: Crack the Code: Mastermind is the Ultimate Logic Puzzle Game, https://ku-biyagames.com/blogs/mechanical-puzzles-blog/crack-the-code-mastermind-is-the-ultimate-logic-puzzle-game, last accessed 2024/03/03
28. Sahana: 9 Advantages and Disadvantages of Joystick, https://www.techquintal.com/ad-vantages-and-disadvantages-of-joystick/, last accessed 2024/03/31

User-centred design and development of a web-based Western Cape substance use assessment tool (WC-SUDAT)

Cameron Worthington[1][0000-0003-2405-8293], Leon Holtzhausen[2][0000-0003-1883-9843], and Michelle Kuttel[1][0000-0003-0554-4632]

[1] Department of Computer Science, University of Cape Town, Rondebosch, 7700, South Africa
[2] Department of Social Work, University of Cape Town, Rondebosch, 7700, South Africa

Abstract. Substance use disorders (SUDs), the uncontrolled use of substances despite harmful consequences, is a significant problem in South Africa, especially in the Western Cape. An important component in the fight against SUDs are questionnaires to assess the risk of an SUD, that are administered by social workers to identify targeted interventions. A web-based questionnaire with automated aggregation of responses can reduce the administrative burden placed on social workers. Here we use a user-centred design approach to build a web-based substance use disorder assessment tool localised to the Western Cape: WC-SUDAT. Our three-phase User Centred Design methodology comprised a first prototype; followed by evaluation of its suitability through a contextual inquiry, a usability test and heuristic evaluation; and then implementation of a final prototype incorporating unanticipated features critical for field use that were identified in the evaluation. This process was effective in generating a final prototype webtool with a dual function as both an SUD assessment tool and an organisational management tool. This deployment-ready prototype is a better fit for the needs of NGOs working with substance abuse disorders than our original conception of the webtool, thus validating a User-Centred design approach.

Keywords: web development, user testing, human computer interaction, user-centred design, substance abuse disorder

1 Introduction

Substance use disorders (SUDs) are endemic in South Africa and contribute to mental, social, and physical health problems [18], most particularly in the Western Cape. NGOs funded by the Western Cape Department of Social Development (DSD) run three SUD treatment programmes: early intervention (EI) to identify and treat at-risk clients before they show symptoms of an SUD; community-based treatment (CBT) to treat an SUD by building a community of support around the client; and aftercare and re-integration (ARI) to help the client adapt to everyday life after treatment.

Screening tools are questionnaires used in EI programmes to identify subjects who may be experiencing, or are at risk of developing, an SUD [1,10,3,6]. A standardised paper-based assessment tool localised to the Western Cape was recently piloted at SUD NGOs in the Western Cape to assess a client's SUD risk level and risk factors (WC-SUDAT) [7]. This is administered by social service professionals in a paper-based format; a web-based screening tool has the potential to improve and accelerate the screening process and lessen the administrative burden on social workers[6,1,17,15,16,8]. Computerised screening also has the potential to be integrated with eHealth records [5,16].

We followed a three phase User Centred Design (UCD) process comprising the build of a first prototype localised web-based version of the WC-SUDAT [7] questionnaire; followed by evaluation of its suitability through a contextual inquiry, a usability test and heuristic evaluation; and then implementation of a final prototype incorporating unanticipated critical features needed for field use that were identified in the evaluation. The participatory design process involved users from two NGOs in the Western Cape. The first prototype was evaluated for functionality and usability with a contextual enquiry, heuristic evaluation and a usability test with the System Usability Scale (SUS), which has been found to be effective in usability evaluations of an eHealth web tool [14]. We incorporated feedback from this process to develop a second prototype tool ready for deployment.

2 Methodology

We followed a UCD process used for development of our prototype, incorporating users into the design process. Human centred design (HCD), as defined by ISO 9241-210:2019, is a design methodology that requires developers to consider all people as potential users and so requires developers to build for a wide range of people. UCD [4] is a more refined version of HCD, but the two terms are often used interchangeably. UCD requires developers to define their user base (which is not all humans, as in HCD) and then build empathy for their users. This can be accomplished using methods such as: a contextual inquiry [26], where the developer interviews end-users to understand their workflows; personas [25], where the developers create imaginary users that represent certain user demographics uncovered by their contextual inquiries; and day-in-the-life studies [27], a quick method in which users sketch their day to help developers understand inefficiencies. Other design methodologies include persuasive design (PD) [29], where developers analyse what could influence their user's behaviour and then build those principles into the design to make it more compelling; and participatory design, where the users are included in the design process from the start of the project by taking part in brainstorming workshops to decide on features and solutions [28]. Development of related eHealth web tools with UCD indicates that focus groups are useful for conducting usability interviews [13,14,19] and unstructured and semi-structured questions allow "digging deeper" to provide valuable insights[24,20,9,14]. Marien et al. suggested that five participants is sufficient, due to the time constraints of conducting lengthy usability assessments [14]. Both qualitative and quantitative data capture in a usability study is important for the quality of the study and the System Usability Scale (SUS) are effective for quantifying the usability of an eHealth web tool [14,24,20,19,9,13].

2.1 Approach

We aim to identify and implement key features in the webtool to enable it to be adopted in the EI programmes for SUD treatment. We identified two categories of potential user of our webtool: either clinicians, who are client-facing, or researchers, who need access to an anonymised database of client records. Our UCD process incorporated two client NGOs: the Cape Town Drug Counselling Centre (CTDCC), a multi-branch NGO in the Western Cape with EI, CBT, and ARI programmes covering the full range of support that is funded by the DSD, and the Knysna Alcohol and Drug Centre (KADC), a single-branch NGO offering EI and ARI programmes.

The webtool was developed in three phases, as follows. Phase One developed a first prototype prototype (V1) that encoded the WC-SUDAT [7] questionnaire. In Phase Two, the suitability of the V1 prototype was evaluated in three ways with users from CTDCC and KADC: a **contextual inquiry** enabled the development team to understand the users through job shadowing; a **usability test** assessed the tool with a set of users; and a **heuristic evaluation** tested interface with trained evaluators. Phase Three addressed fundamental issues raised in the evaluation phased implementing core necessary features to create a beta build of the webtool (prototype V2) that is fit for purpose and WC-SUDAT ready for deployment.

3 Phase One: Prototype V1

The basic functionality implemented in prototype V1 allows clinicians to create clients, save their details, and administer assessments from which SUD risk levels are calculated. Once the assessment is administered to a client, the clinician can interpret the risk visualisation and determine the path forward for their client. If the clinician needs to focus on a specific set of the clients answers, they can view the assessment again along with the notes they may have taken during the assessment. This use case is represented in the core features: user accounts; organisational-based access for clinicians; client creation and management; implementation of the WC-SUDAT assessment; visualisation of the client's SUD risk levels grouped by risk factors for each completed assessment; assessment history and clinician comments; and a feedback mechanism for communication with the developers.

There are four pages accessible once a user is logged in: the home page for clinicians for creation and editing of new and existing clients; a page displaying the results of an assessment, with risk factors and a client's risk levels; the page with the WC-SUDAT assessment where questions are read to a client by a clinician who records the client's answers; and the researcher page presenting all the anonymised client answers.

The *user accounts* feature enables user registration and, once signed in, allows access to all the other functionality of the tool. The feature is implemented with the Django app textitAccounts that securely stores user's passwords with usernames (unique identifier). We extended the model to enable the user's unique identifier to be the email address. Rudimentary log in and registration pages were built using static html forms. Lastly, we configured the Django admin portal to do basic create, read, update and delete (CRUD) functions on the users, including password reset, since no user flow for this had been implemented. The user model has a unique South African Council for Social Service Professions (SACSSP) registration number attribute that is used to retrieve clients, as well as a supervisor's SACSSP registration number for the cases of student social workers that may have a supervisor who must able to see their clients.

The *organisational-based access for clinicians* feature allows users in the same organisation to view and administer assessments to one another's clients so that clients can be attended to in the event that their social worker is unavailable. The implementation attached an organisation model to the user object. The data the user can access is filtered based on SACSSP registration number and on the organisation that they are a part of. Organisations can be managed in the admin portal.

The *client creation and management* feature allows social workers to register and manage their clients personal details. The client model has all the attributes that WC-SUDAT requires for data analysis. Clients are assigned only one registered social worker, however, they are visible to all social workers in their social worker's organisation and to their social worker's supervisor.

The *implementation of the assessment tool* allows an administrator to create or edit an assessment via the Django admin portal. Once created, a user can administer the assessment to a client. The WC-SUDAT assessment consists of long form answers, Likert scale questions and yes/no questions. The Likert and yes/no questions can unlock follow up questions and have risk calculations based on the answers. For the implementation, the assessment is generalised into sections, subsections and two question types: text answer questions, which have a text field as the answer input; and choice answer questions, which can be assigned choices that allow these questions to be yes/no or Likert questions. The choice questions and choice answers (the objects that can be assigned as answers to choice questions) can be assigned risk factors and risk values, respectively. Risk factors are grouped into risk categories for reporting purposes. Sections and subsections can be assigned risk thresholds that determine the total categorical risk a client must have to unlock that subsection. Each section and subsection also have a text field that allows clinician notes to be taken during the assessment. This implementation gives enough flexibility to completely digitise the WC-SUDAT assessment tool.

The *visualisation of the client's SUD risk levels grouped by risk factors for each completed assessment* feature allows for a visual representation of the clients risk profile as calculated from a completed WC-SUDAT assessment. The visualisations displays all the risk categories and their factors that have been configured in the admin portal. A toggle allows selected risk factors to not be displayed, primarily if they are used for the logic of which questions to show and do not have application for the social worker. A client can have multiple assessments so the visualisation displays the result of the selected assessment as well as the average result. The implementation makes use of a Javascript graph library to visualise the risk factors. Each risk factor has description that can be configured in the admin portal to provide context and explanation to the user.

The *assessment history and clinician comments* feature is a non editable replica of the assessment. It displays all the answers to the assessment. Although the clients answer cannot be edited, the clinician notes are editable in this view.

The *feedback mechanism* feature allows the user to provide feedback to the developer team, it was intended to be used throughout the evaluation process. It saves the feedback in the admin portal.

At this point the tool was considered a minimum viable product, however the following four features were added before conducting the evaluations in Phase Two.

The *email password reset and user roles approval* feature allows all users to reset their password and confirm their email addresses, and organisation administrators to approve new users, granting them access to the WC-SUDAT system according to their role. This functions via email: the system emails the user a link to perform one of the three tasks (email confirmation, user approval, or password reset). An email confirmation link is emailed to the user after registration, the user approval link is emailed to the organisation administrator after a user registers, and the password reset link is emailed after clicking "Forgot password" on the login page and following the prompts. A new user cannot sign in until they have confirmed their email and cannot access the tool until they are approved by the organisation administrator.

The *branch-based user permissions* feature caters for multi-branch organisations by allowing branch-wide client access to all users at the branch and is important for allowing social workers all social workers in a branch to access all clients. The branch model has a parent organisation, address and branch manager as fields. The user and organisation models have an assigned branch and head office as fields, respectively. The system filters client objects by matching the logged-in user's branch to the client object's social worker's branch. This filter happens every time a URL that requests client data is rendered to ensure that there is no unauthorised access.

The *Google Places API for client location* feature implemented in the client creation form enables WC-SUDAT to search Google Maps for location data. This feature prevents errors when capturing a client's location and allows for a range of precision in capturing location. This is important to standardise location data for clients living in informal settlements. Organisations have varying policies on capturing the client's location: some feel that street addresses are too high precision and could be used to arrest clients, with Google Places they can choose how specific they want to be.

The *answer-dependent question access* feature allows dynamic control of which questions appear in the client assessment, dependent on the answers to previous questions. This was done by implementing a question-answer requirement model to allow combinations of questions and answers to be requirements for other questions to be shown.

The prototype was built using Python to implement the controller and model of the Model View Controller (MVC) architecture (Model Template View in Django) and JavaScript, HTML and CSS to implement the View. All prototypes were developed with the Agile methodology, which relies on multiple iterations of development and testing [2].

The prototype was *deployed* on the Department of Computer Science's servers at the University of Cape Town with the URL wcsudat.cs.uct.ac.za.

4 Phase Two: Prototype Evaluation

We followed a UCD approach to evaluate the first prototype of WC-SUDAT. using contextual enquiries to gain empathy for the user and understand the WC SUD NGO processes; a usability test to evaluate how well the webtool performs in the field; and heuristic evaluations to identify usability issues.

4.1 Contextual Enquiry

We conducted two **contextual inquiries** to identify the overlap between two SUD organisations' requirements for a webtool. As we already had a webtool prototype, the gaps between the users requirements and the functionality of our tool were more easily identified. The first inquiry was conducted at the Cape Town Drug Counselling Centre (CTDCC) to understand the organisational processes of an SUD clinic. The inquiry followed the director of the organisation and the head social worker through the client intake process and the compilation of quarterly reports for the DSD, to understand the entire paper trail of a client from intake to quarterly report. A second smaller contextual inquiry consisted of an unstructured interview conducted over Zoom with the Knysna Alcohol and Drug Centre (KADC).

These contextual enquiries identified the following three key requirements for a webtool.

Reduced assessment time. SUD organisations administer multiple assessments are administered to each client, which is time-consuming. Some of the assessment's questions overlap, which wastes time in repeating the answers. Self-administered assessments are not desirable as the KADC said that interaction with the client during an assessment helps to inform the diagnosis. Therefore assessment questionnaires need to be as short as possible, and the most important questions should be answered first.

Reduced administration for social workers. A client's file includes an assessment, a treatment plan, counselling notes, summaries of interactions with parents or employers, and the drug tests conducted. Files are accessed by a client's social worker, other social workers, sessional staff (art therapist, doctor and psychiatrists), and the DSD. In addition to the assessments, for each client social workers must scan the paper documents, including handwritten observations and professional opinions, and count the number of drug tests (positive and negative) per client as important information for both the client's file and for generating reports. Consolidated quarterly reports are sent to the DSD and the City of Cape Town. Generating quarterly reports is a complex process with multiple steps (Figure 1). The DSD funds three programs: Community Based Treatment (CBT), Aftercare and Re-integration (ARI), and Early Intervention (EI). Social workers in each branch of an organisation capture client's data and the interventions on an Excel spreadsheet for each of the programmes (CBT, ARI, and EI) and send them to the director. The director then create a consolidated Excel sheet for that branch (CBT-branchX, ARI-branchX, and EI-branchX), resulting in three files per branch and also compiles narrative progress reports for each program: for every branch six documents are sent to the DSD. Twice a year, each branch has an on-site visit during which the client's files are inspected. An automated system which reduces this administrative burden, particularly for report generation, would be very valuable.

Adherance to operational constraints. NGOs need to ensure that every social worker in a branch can access each other's clients. Currently, everything is filled out on paper and kept in a file. Replicating this system would lead to better uptake with the social workers. In addition, POPIA [South Africa. Protection of Personal Information Act of 2013] and client confidentiality must be ensured. Currently all client's data must be hosted on-site at the branch head office.

4.2 Usability Test

The usability test was conducted with four social workers from the CTDCC and four from the KADC organisations in two parts: a SUS questionnaire [12,11] (a standardised 10-question questionnaire to calculate a usability score out of 100), followed by semi-structured interviews. The tests took place at organisation's offices, which are the real-world locations where the tools would be used. Of note is that both organisations only recently purchased laptops for their staff; hence the user's may be unfamiliar with the technology. Social workers were given a week to use the webtool. They were required to add a client to the tool and complete an assessment for that client. This task required them to perform multiple sub-tasks: registering a user account; signing in; registering a new client; administering an assessment; and analysing the assessment feedback for that client. After using the tool, the users completed a SUS questionnaire and participated in semi-structured interviews, which further explored the usability of the tool and any feature changes they would need for the tool to be adopted as part of their processes. There were some significant real-world challenges that cause significant delays in completing the usability tests: taxi strikes prevented

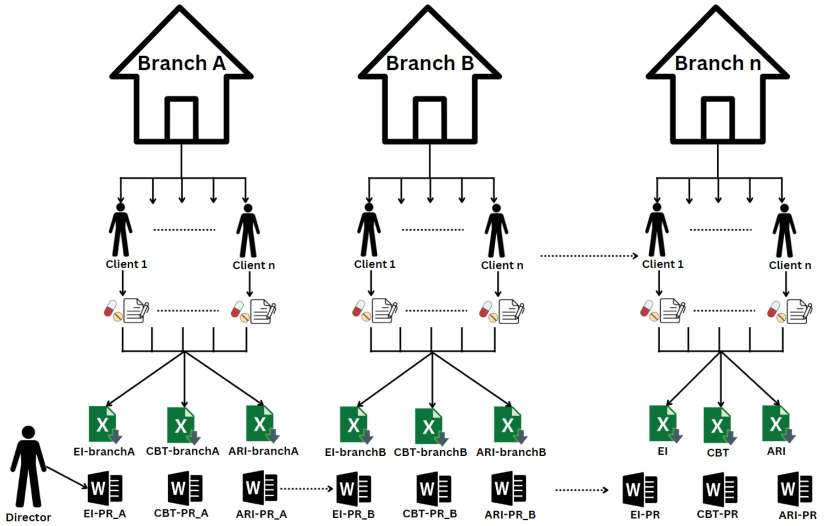


Fig. 1. The process of compiling quarterly reports for a multi-branch organisation

organisations from operating, a death at one of the organisations, and one of the branches of a organisation burnt down.

During the interview with the CTDCC social workers, it became apparent that their usability concerns were primarily with the questions in the assessment rather than the interface. To mitigate this users at CTDCC completed two SUS tests: for the first test they were asked to evaluate the usability of the assessment itself instead of the interface; a second test was then completed focussing only on the usability of the interface. The first test had an average usability score of $51.3 \pm 6.3\%$ within the 10th percentile of interface usability [12]. The SUS questionnaire focussing only on the interface had an average score of $72.5 \pm 5.9\%$, within the 64th percentile of interfaces. One outlier of 62.5 (included in the mean calculation) was obtained from P1, the oldest in the group of participants.

The interviews highlighted the issues in more depth, as follows.

The sign-up page was unusable. This test highlighted poor implementation of the sign-up process in prototype V1. One participant said "It was terrible" even after email support from the development team. The main issue was the lack of useful error messages: different errors used the same message, a participant said, "It wasn't very specific".

Overview on the home page. Although users said that interaction with the home was "fairly straightforward" (it was easy to create a client and the jump to the client page once a new client is registered was "actually cool") they said that they would prefer an overview page on the home page. One participant mentioned that it could display the status of their clients saying, "These are my CBT clients, these are my early intervention clients, and so on".

Confusion with the assessment. Participants were not familiar with the assessment implemented in V1. Some found answering the questions confusing ("you had to first think yes I am not able to and then no I am not able to"; "I thought, what's going on?"). Participants mentioned an overlap in questions and that different sections would "ask the same sort of question in a different way". Participants felt that questions were closed-ended and made a "barrier between client and social worker". They were concerned that it "shuts down a flow in conversation". One participant mentioned that they, "walked away from my assessment, not knowing much [about the client]" and that the assessment was "very surface level". They felt that some questions played into the psychological defences of SUD. These were things such as blaming it on external factors (in the sections that ask how much your community/family affected your SUD). One participant said, "Someone could say, this is my girlfriend's fault or my sister's". But then the participant considered that "this is more of a critique on all assessment tools".

Resistance to change of assessment. One participant said, "We've been working with our own assessment for so long that I gravitate towards accepting that." They mentioned that the WC-SUDAT questions are "very specific" and that their questions are "a lot broader" but have fewer sections. Their questionnaires do not look into school-based factors. They agree that new questions are important and that "we need fresh eyes on it" because "we can't respond to a growing and a different external environment doing the same thing".

Assessment is too long. Participants unanimously agreed that the questionnaire was too long ("took 50 minutes to close to an hour") and would take up important conversation time with A client. The completion time is an issue because the webtool complements existing assessments since it does not ask all the questions that the social workers need e.g. "we did not know what [substances] they were using".

Risk assessment available beforehand The risk breakdown needs to be accessible before talking to the client. If the social worker could see the risk factors before engaging with the client then they could use them to inform their conversation with the client. One participant said that "in an ideal world we could say to a client [before coming to the appointment], log into our app on the website and fill out the assessment". They were only concerned about the practicality of this. One participant said that "the only issue is the resources".

Concerns with digitisation of their assessments and processes. Social workers are used to writing everything down. One participant said, "I'm so used to writing by hand, typing it out may be a little bit challenging". Another issue was that clients usually complete pre-appointment assessments on paper. To digitise, the clients would need access to a computer to complete the assessments. One participant pointed out that the place to make notes about a client should resemble the paper-based assessment tool and should come after the section of questions, and not on the side. When asked if they use a digital calendar, a participant said, "A digital calendar would be great". Currently, they keep track of everything on paper-based diaries.

Other than the sign-up page, the interview comments about the usability of the interface were all positive. One participant said that the "system itself seemed to flow" when asked about the interface. One participant said that the tool *"wasn't a scary tool to use"*. This statement is worth noting, as the same participant does not "make online purchases because they are too complicated".

4.3 Heuristic evaluation

For the heuristic evaluation, the evaluators were given heuristic checklist to critique the design comprising all 10 of Nielsen's heuristics [21] and five additional guidelines (Table 2). Each problem identified by the evaluators was related to a heuristic and given a severity rating (Table 1) out of four, which indicates the urgency of fixing the problem. The evaluators and developer then discussed possible solutions to the problem. This is an effective way to find, prioritise and solve usability problems [23]. We used two evaluators who had previously completed heuristic evaluation courses.

Table 1. Severity ratings for each usability issue for the heuristic evaluation as per the Nielsen Norman's group "Severity Ratings for Usability Problems"[22,30].

Severity Rating	Explanation
1	Cosmetic problem
2	Low-priority usability problem
3	High-priority usability problem
4	Usability catastrophe (imperative to fix)

The evaluators determined that V1 of the prototype covered all but two of the selected heuristics (86%), with only *help and documentation* and *structure of information* not covered adequately. The insufficient *help and documentation* was rated a three, a high-priority usability problem (Table 1). Evaluators said that the tool felt overwhelming to a first-time user. A potential solution is an on-screen walk-through or tutorial on the first login, which would subsequently be accessible through a help button.

The home page of the tool was overwhelming due to the client creation form being immediately visible, which gives a new user too much information too soon. This is in contravention of the the *structure of information* and *aesthetic and minimalist design* heuristics. Evaluators suggested replacing the form with an overview of the status of the user's clients and providing a button to show the client creation form *on demand*. In addition, we could apply *flexibility and efficiency of use* heuristic and allow the user to choose the default view of the homepage. The severity of the issue was rated a three, a high-priority usability problem.

Table 2. Nielsen's ten and five additional heuristics used for the heuristic evaluation[21]

Heuristic	**Description**
Visibility of system status	Reasonable and timely feedback to inform the user what is happening.
Match between the system and the real world	Avoid unfamiliar terms or processes by emulating the user's environment.
User control and freedom	Users need to leave unwanted states easily and support undo and redo.
Consistency and standards	Ensure the system is consistent and follows platform conventions.
Error prevention	Design for error prevention and present useful error messages if you cannot avoid the error.
Recognition rather than recall	Make relevant actions and information visible to reduce memory load.
Flexibility and efficiency of use	Cater to both inexperienced and experienced users by allowing users to tailor frequent actions.
Aesthetic and minimalist design	Irrelevant or rarely needed information should be avoided.
Recognize, diagnose, and recover from errors	Error messages should be intuitive and plain while also providing quick recovery options.
Help and documentation	Provide natural help and documentation to the user
Navigation	Provide navigation aids (search functionality) and give feedback about where the user is
Use of modes	The system caters for a variety of modes
Structure of information	Information is presented simply and understandably
Enjoyment	The system is fun and satisfying to use
Extraordinary users	Cater for a wide variety of users, including those with disabilities

A smaller usability problem was the lack of an overview for the users which falls under *system status* with a rating of two, a low priority. This would be investigated in the usability test to see what information the users want in an overview.

A problem with the auto-scroll, the system that automatically moves the user to the next question, was identified under *user control and freedom*. When the auto-scroll of an assessment is on (it can be toggled) and the user skips a section and continues further below, the screen "whips" back to the next incomplete question in a jarring motion. This is usually unintentional on the user's part. The severity was rated a three, a high priority, with the solution being to make the auto-scroll never jump to previous sections and to just scroll to the next incomplete question relative to the user's position in the assessment.

A final usability suggestion was made, under *flexibility and efficiency of use*, that the search bar on the home page should filter using more than just the client's name. The search could also filter clients based on other details, such as file number or contact details. The severity was rated a two, since it would be a useful feature but is not a usability problem.

5 Phase Three: Prototype V2

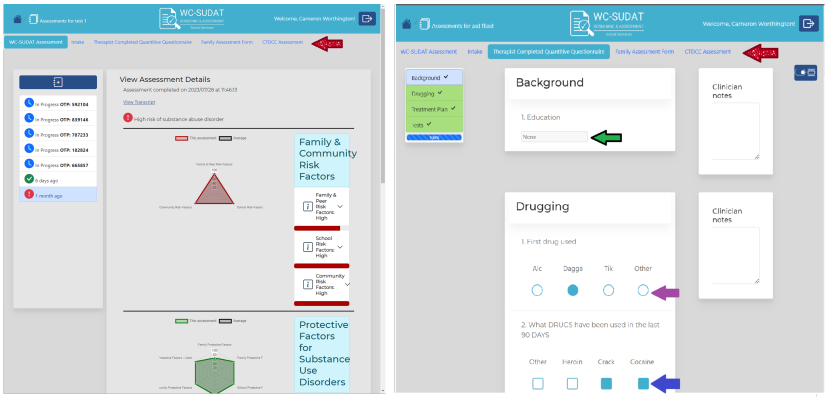
In accordance with the findings of the contextual inquiry, usability test and heuristic evaluation, prototype V2 was reformulated to have a dual purpose, operating as both an SUD assessment tool and an organisational management tool.

Prototype V2 enables digitisation of any forms that a SUD organisation would use. The tool incorporates the idea of a test suite, allowing organisations to add multiple assessments or forms for their clinicians to use in conjunction with one another. With this much expanded assessment and information storing system, all the data for the DSD can be captured. The intention is to ultimately make generating quarterly reports a seamless process (not yet implemented). Upgrades to the WC-SUDAT assessment now enable clients to complete it on their own using a OTP to access their unique assessment. This should speed up the on boarding process for a new client. One caveat to the tool is that the risk calculation is only applicable to the WC-SUDAT assessment. This is to incentivise the use of WC-SUDAT as there are benefits to standardising an assessment tool of this nature.

Most importantly, the beta build digitises the paper-based processes of Western Cape SUD NGOs, a central theme highlighted by the contextual inquiry to allow for automation of processes (such as DSD data collection).

Four key features were implemented: a *multiple assessments feature*, which allows multiple assessments to be added to the tool; *sign-up page validation*, to provide helpful error messages to guide users through the sign-up process; *one-time pin (OTP) based assessment access*, to allow social workers to generate assessments that a client can complete without logging in to the tool; and *DSD data capture*, to add all the client data capture required by the DSD. All features were tested manually or through unit tests.

V2 allows for *multiple assessments* or forms to be administered: the prototype can digitise any assessment or form using the same style as the original questionnaire (Fig. 2). This is a fundamental change to the original prototype. An assessment comprises sections, which have subsections containing three question types: text answer questions (which have text input as their answer, Fig. 2 pink arrow); choice answer questions (which have choices from which a user can select one or multiple answers, Fig, 2 blue arrow); and client detail questions (which take fields of a client object and insert them into an assessment, Fig, 2 green arrow with black border). Assessments and forms appear as tabs on the client page which a user can switch between (Figure 2 maroon spotted arrow). Assessments are only displayed to the users of the organisation that created them. The risk report was updated to work with multiple assessments. In addition, every client may one of each assessment; the assessment questions and client answers can be updated at any time, assessments may log client information that is already captured, and assessments may require multiple choice answers per question.



View Assessment Details

Assessment completed on 2023/07/28 at 11:46:13

View Transcript

High risk of substance abuse disorder

Background

1. Education

None

Drugging

1. First drug used

○ Alc ● Dagga ○ Tik ○ Other

2. What DRUGS have been used in the last 90 DAYS

□ Other □ Heroin ■ Crack ■ Cocaine

Fig. 2. Prototype V2 client page dynamically loads all the assessments that an organisation has digitised. The left image is the original assessment, the right image one of the digitised assessments of CTDCC. There are three question types: text answer questions (pink arrow); choice answer questions (blue arrow); and client detail questions (green arrow with black border). Assessments and forms appear as tabs on the client page which a user can switch between (maroon spotted arrow)

The sign-up page in prototype V2 validates all data fields with the level of detail in error messages brought in line with other websites (Google and Facebook) and occasionally provides more detail (e.g. the password field tells the user exactly what character types they require to make the password secure). The page is organised to avoid errors, for example, the dropdown for selecting an organisation filters the branches depending on which organisation the user selected. A one-time pin (OTP) based assessment access feature was added to enable a social worker to generate an assessment that a client can access with a 6-digit unique assessment code (or OTP) without logging in.

All features were implemented using client page HTML, CSS and Javascript.

6 Discussion

Our User Centred Design methodology had three phases: building a a first throw-away prototype of a webtool questionnaire as a straw man in Phase One, and then subsequent contextual enquiry and evaluation with a quantitative usability test and qualitative heuristic evaluation in Phase Two. As we already had a webtool prototype, the gaps between the users requirements and the functionality of our tool were more easily identified in the contextual enquiry than if this had been performed before Phase One. This approach highlighted fundamental issues with the both the focus and implementation of the first prototype, which were addressed in Phase Three of our second implementation which is fit for purpose and ready for deployment.

A primary issue raised in both the contextual enquiry and the usability test is that prototype V1 did not have sufficiently broad and useful functionality: SUD organisations do not want another standalone assessment tool. Because social workers spend a large amount of time on assessments and administrative tasks, a webtool must function not only as an assessment tool, but also a client and organisational management tool. This dual purpose is critical for the uptake of the webtool by NGOs.

We addressed this in the second prototype reformulating the tool to to have a dual purpose: an SUD assessment tool and an organisational management tool. We added extensive additional functionality to digitise the paper-based processes of Western Cape SUD NGOs and so allow for automation of processes such as data collection for the DSD which will be useful for the generation of quarterly reports by NGOs. . We also implemented other desirable features included validation on the user sign-up page; and OTP-based assessment access by clients..

Another key finding is that users were unhappy with the format of the questions and length of the questionnaire used for assessment, to the extent that this impeded assessment of the tool's usability. Although the interface had an average SUS score of 72.5 ± 5.9% and is more usable than 64 % of interfaces currently in use [12], we found that users were assessing the questions in the assessment rather than the usability of the tool.

We addressed this in the second prototype by allowing multiple and alternative assessments to be administered in additional to the original assessment. V2 of the prototype can digitise any assessment or form using the same style as the original questionnaire.

Deployment of prototype V2 will require a distributed database to store the organisation's client information on-site. However, before deployment, the issue of POPIA and client confidentiality must be addressed. The constraint is that private client data (name, surname and ID number) must be stored on-site at the organisation. This could be done by deploying an instance of the database onsite which stores the client data only. This would require a small computer at every organisation's head office.

7 Conclusions

Our three-phase User Centred Design methodology was effective in generating a final prototype webtool with a dual purpose as an SUD assessment tool and an organisational management tool. Although the prototype developed fulfils our aims, there are a number of possible future additions to the tool. Most beneficial would be to integrate the data captured with the Department of Social developments quarterly reporting processes. This is a complex task that will require an extensive further UCD process, and hence is outside the scope of this project.

The development of our WC-SUDAT webtool is a case study in the value of using UCD to develop effective software for the public sector in South Africa. Our final deployment-ready prototype is a better fit for the needs of NGOs working with substance abuse disorders that our original webtool, thus validating the User-Centred design approach.

References

1. Adam, A., Schwartz, R.P., Wu, L.T., Subramaniam, G., Laska, E., Sharma, G., Mili, S., McNeely, J.: Electronic self-administered screening for substance use in adult primary care patients: feasibility and acceptability of the tobacco, alcohol, prescription medication, and other substance use (myTAPS) screening tool. Addiction Science & Clinical Practice **14**, 1–12 (October 2019). https://doi.org/https://doi.org/10.1186/s13722-019-0167-z
2. Fowler, M., Highsmith, J., et al.: The Agile manifesto. Software development **9**(8), 28–35 (2001)
3. Gryczynski, J., Kelly, S.M., Mitchell, S.G., Kirk, A., O'Grady, K.E., Schwartz, R.P.: Validation and performance of the alcohol, smoking and substance involvement screening test (ASSIST) among adolescent primary care patients. Addiction **110**(2), 240–247 (October 2015). https://doi.org/https://doi.org/10.1111/add.12767
4. Gulliksen, J., Göransson, B., Boivie, I., Blomkvist, S., Persson, J., Åsa Cajander: Key principles for user-centred systems design. Behaviour & Information Technology **22**(6), 397–409 (2003). https://doi.org/10.1080/01449290310001624329
5. Harris, S.K., Knight, J.R.: Putting the screen in screening: technology-based alcohol screening and brief interventions in medical settings. Alcohol Research: Current Reviews **36**(1), 63—79 (2014)
6. Harris, S.K., Knight, Jr, J.R., Van Hook, S., Sherritt, L., L. Brooks, T., Kulig, J.W., A. Nordt, C., Saitz, R.: Adolescent substance use screening in primary care: Validity of computer self-administered versus clinician-administered screening. Substance abuse **37**(1), 197–203 (2016). https://doi.org/https://doi.org/10.1080/08897077.2015.1014615
7. Holtzhausen, L.: The Western Cape Substance Use Disorder Assessment Tool (2023), (Unpublished)
8. Inman, D., El-Mallakh, P., Jensen, L., Ossege, J., Scott, L.: Addressing substance use in adolescents: Screening, brief intervention, and referral to treatment. The Journal for Nurse Practitioners **16**(1), 69–73 (2020). https://doi.org/https://doi.org/10.1016/j.nurpra.2019.10.004
9. Kawasaki, S., Mills-Huffnagle, S., Aydinoglo, N., Maxin, H., Nunes, E.: Patient- and provider-reported experiences of a mobile novel digital therapeutic in people with opioid use disorder (reSET-O): Feasibility and acceptability study. JMIR Form Res **6**(3), e33073 (March 2022). https://doi.org/10.2196/33073
10. Knight, J.R., Sherritt, L., Harris, S.K., Gates, E.C., Chang, G.: Validity of brief alcohol screening tests among adolescents: a comparison of the AUDIT, POSIT, CAGE, and CRAFFT. Alcoholism: Clinical and experimental research **27**(1), 67–73 (2003)

11. Kortum, P.T., Bangor, A.: Usability ratings for everyday products measured with the System Usability Scale. International Journal of Human–Computer Interaction **29**(2), 67–76 (2013). https://doi.org/10.1080/10447318.2012.681221
12. Lewis, J., Sauro, J.: Can I leave this one out? The effect of dropping an item from the SUS. Journal of Usability Studies **13**, 38–46 (November 2017)
13. Lowe, C., Browne, M., Marsh, W., Morrissey, D.: Usability testing of a digital assessment routing tool for musculoskeletal disorders: Iterative, convergent mixed methods study. J Med Internet Res **24**(8), e38352 (August 2022). https://doi.org/10.2196/38352, http://www.ncbi.nlm.nih.gov/pubmed/36040787
14. Marien, S., Legrand, D., Ramdoyal, R., Nsenga, J., Ospina, G., Ramon, V., Spinewine, A.: A user-centered design and usability testing of a web-based medication reconciliation application integrated in an eHealth network. International Journal of Medical Informatics **126**, 138–146 (2019). https://doi.org/https://doi.org/10.1016/j.ijmedinf.2019.03.013, https://www.sciencedirect.com/science/article/pii/S1386505618301151
15. McNeely, J., Cleland, C.M., Strauss, S.M., Palamar, J.J., Rotrosen, J., Saitz, R.: Validation of self-administered single-item screening questions (SISQs) for unhealthy alcohol and drug use in primary care patients. Journal of general internal medicine **30**, 1757–1764 (May 2015). https://doi.org/https://doi.org/10.1007/s11606-015-3391-6
16. McNeely, J., Strauss, S.M., Rotrosen, J., Ramautar, A., Gourevitch, M.N.: Validation of an audio computer-assisted self-interview (ACASI) version of the alcohol, smoking and substance involvement screening test (assist) in primary care patients. Addiction **111**(2), 233–244 (February 2016). https://doi.org/https://doi.org/10.1111/add.13165
17. McNeely, J., Strauss, S.M., Saitz, R., Cleland, C.M., Palamar, J.J., Rotrosen, J., Gourevitch, M.N.: A brief patient self-administered substance use screening tool for primary care: two-site validation study of the substance use brief screen (SUBS). The American journal of medicine **128**(7), 784–e9 (July 2015). https://doi.org/https://doi.org/10.1016/j.amjmed.2015.02.007
18. Myers, B., Koch, J.R., Johnson, K., Harker, N.: Factors associated with patient-reported experiences and outcomes of substance use disorder treatment in Cape Town, South Africa. Addiction Science & Clinical Practice **17**(1), 8 (February 2022). https://doi.org/https://doi.org/10.1186/s13722-022-00289-3
19. Neubeck, L., Coorey, G., Peiris, D., Mulley, J., Heeley, E., Hersch, F., Redfern, J.: Development of an integrated e-health tool for people with, or at high risk of, cardiovascular disease: The consumer navigation of electronic cardiovascular tools (CONNECT) web application. International Journal of Medical Informatics **96**, 24–37 (December 2016). https://doi.org/https://doi.org/10.1016/j.ijmedinf.2016.01.009
20. Neville, C., Da Costa, D., Rochon, M., Peschken, C.A., Pineau, C.A., Bernatsky, S., Keeling, S., Avina-Zubieta, A., Lye, E., Eng, D., Fortin, P.R.: Development of the lupus interactive navigator as an empowering web-based ehealth tool to facilitate lupus management: Users perspectives on usability and acceptability. JMIR Res Protoc **5**(2), e44 (May 2016). https://doi.org/10.2196/resprot.4219
21. Nielsen, J.: Enhancing the explanatory power of usability heuristics. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp. 152–158. CHI '94, Association for Computing Machinery, New York, NY, USA (1994). https://doi.org/10.1145/191666.191729, https://doi.org/10.1145/191666.191729

22. Nielsen, J.: Usability inspection methods. In: Conference companion on Human factors in computing systems. pp. 413–414 (1994)
23. Nielsen, J.: How to conduct a heuristic evaluation. Nielsen Norman Group **1**(1), 8 (1995)
24. Ospina-Pinillos, L., Davenport, T.A., Ricci, C.S., Milton, A.C., Scott, E.M., Hickie, I.B.: Developing a mental health eclinic to improve access to and quality of mental health care for young people: Using participatory design as research methodologies. J Med Internet Res **20**(5), e188 (May 2018). https://doi.org/10.2196/jmir.9716, https://doi.org/10.2196/jmir.9716
25. Pruitt, J., Grudin, J.: Personas: Practice and theory. In: Proceedings of the 2003 Conference on Designing for User Experiences. p. 1âĂŞ15. DUX '03, Association for Computing Machinery, New York, NY, USA (2003). https://doi.org/10.1145/997078.997089, https://doi.org/10.1145/997078.997089
26. Raven, M.E., Flanders, A.: Using contextual inquiry to learn about your audiences. SIGDOC Asterisk J. Comput. Doc. **20**(1), 1âĂŞ13 (feb 1996). https://doi.org/10.1145/227614.227615, https://doi.org/10.1145/227614.227615
27. Samaroo, R., Brown, J.M., Biddle, R., Greenspan, S.: The day-in-the-life scenario: A technique for capturing user experience in complex work environments. In: 2013 10th International Conference and Expo on Emerging Technologies for a Smarter World (CEWIT). pp. 1–7 (2013). https://doi.org/10.1109/CEWIT.2013.6713761
28. Spinuzzi, C.: The methodology of participatory design. Technical communication **52**(2), 163–174 (2005)
29. Torning, K., Oinas-Kukkonen, H.: Persuasive system design: State of the art and future directions. In: Proceedings of the 4th International Conference on Persuasive Technology. Persuasive '09, Association for Computing Machinery, New York, NY, USA (2009). https://doi.org/10.1145/1541948.1541989, https://doi.org/10.1145/1541948.1541989
30. Zhang, J., Johnson, T.R., Patel, V.L., Paige, D.L., Kubose, T.: Using usability heuristics to evaluate patient safety of medical devices. Journal of Biomedical Informatics **36**(1), 23–30 (2003). https://doi.org/https://doi.org/10.1016/S1532-0464(03)00060-1, https://www.sciencedirect.com/science/article/pii/S1532046403000601

A WC-SUDAT code repository

The repository for the project code can be found at this url https://gitlab.cs.uct.ac.za/wrtcam003/wc-sudat

Part II

Information Systems Track

Constraints and Outcomes of Giving Free Wi-Fi Access to Learners: A South African Case Study.

Kevin Birtles[1], Zane Davids[1,2] [0009-0008-2760-5567] and Lisa Seymour[1,2] [0000-0001-6704-0021]

[1] Department of Information Systems, University of Cape Town, Cape Town, South Africa
BRTKEV007@myuct.ac.za, mo.davids@uct.ac.za,
lisa.seymour@uct.ac.za
[2] CITANDA, University of Cape Town, Cape Town, South Africa

Abstract. We cannot imagine a world without the Internet both within our personal life and for educational purposes. In South Africa, the government's "Smart Classroom project" introduced Wi-Fi in schools. However, its adoption by learners has been slow which has been attributed to school management concerns. This research aimed to describe the constraints and perceived outcomes of giving free Wi-Fi access to learners to understand the slow adoption. This qualitative multiple case study in five Western Cape schools used a hybrid inductive-deductive analysis of interviews and documents. The study identified nine constraints preventing learners accessing the Wi- Fi, dominated by school restrictive policies and lack of device control. The study also highlighted both positive and negative outcomes for learners using the school's Wi-Fi. Game-based learning in classrooms and accessing online educational content are some positive outcomes. Negative outcomes are being distracted from learning, cyberbullying, accessing pornography and internet abuse. This study will be of particular interest to education organizations and government decision makers highlighting areas of concerns amongst school management when providing free Wi-Fi access to learners. Finally, the paper suggests that the Western Cape Education Department (WCED) considers increasing their financial investments into procuring additional bandwidth, additional teacher training, and smart applications to help manage personal devices of learners and staff.

Keywords: Wi-Fi Access, Schools, Learners, Constraints, Outcomes.

1 Introduction

Many governments have realized the potential of the internet to boost economic growth across the globe and are willing to subsidize the rollout of Broadband [1]. In 2015 the South African (SA) Western Cape Education Department (WCED) announced that they will be investing in Local Area Networks (LANs) in schools. Wireless distributed network (Wi-Fi) was installed at schools giving access to both learners and staff [2]. In 2022, a total of 752 primary, high, combined, and inclusive schools had received LANs [3]. Yet Wi-Fi use by learners was less than expected. Katsidzira and Seymour [4] identified some constraints in the use of Wi-Fi at schools from a teacher and pupils' perspective and noted that school management had a constraining impact as well. Hence our study chose to focus on a school management’s perspective and asks the question: What are the constraints and outcomes of giving free Wi-Fi access to learners in schools from a school management perspective? The structure of this document begins with a review of the literature on Wi-Fi in education and the formulation of the study's hypothesis. The document then moves on to describe the case study method employed. Subsequently new and validated factors are discussed, the final framework is described, and we conclude.

2 Literature Review

Wi-Fi is an over-the-air link between a wireless client and a base station usually known an access point (AP). Essentially, it gives you the freedom to walk around indoors and outdoors without having to connect your device to the Internet via wires; however, you must be within range of an AP [5-6].

2.1 Outcomes of the Internet for Learners

From a positive perspective, digital technologies combined with the internet enable both educators and learners to benefit from connected classrooms [7]. The internet is the easiest and quickest way of finding accurate necessary information meeting users' needs [8]. Online learning is also stated to be engaging, accessible, and contextualized with the use of audio-video content, activities, a virtual scenario, live interaction with educational professionals, simulations, models, graphics, animations, quizzes, games, and e-notes [9-10]. Furthermore game-based learning can influence the cognitive, emotional, behavioral, motivational, and social elements of learners' involvement in academic fields [11]. As the internet is becoming more accessible via smartphones, learners have used it for collaboration, communication, entertainment, and leisure [12-13]. Integrating mobile technology in the classroom is stated to equip learners for future careers [14]. Hence the Internet is seen as an important educational supplement and means of expanding perspectives [15].

Yet the Internet can negatively impact learners. They can become addicted to internet use, which has a negative outcome on learning [8]. Internet abuse by learners is a persistent problem [16-17]. Learners can also hurt and inflict pain on each other via cyberbullying. A 2015 study of grade 10-12 learners in Western Cape schools, showed that 47% of respondents had been victims of cyberbullying, while 60% admitted they were guilty of cyberbullying [18].

2.2 School Management and Constraints on School Internet Access

School principals need to prepare teachers to use ICT in the classroom [19]. Yet they don’t always understand the potential of smart schools or how to implement them [20]. ICT integration in school can also be adversely affected by personal attitudes, perceptions, and motivations of principals [21]. In addition, teachers need to educate students preparing them for the technological world of the future [22]. Yet successful ICT use in education is influenced by teachers' attitude, skill, and experience with ICT [23]. They need to feel in control and confident when using technology to teach learners, yet many teachers lack the necessary skills [24]. Other constraints constraining access are now reviewed.

Insufficient Devices, Inadequate Connectivity and Technical Support. Many schools lack ICT devices to connect online [25-26]. Furthermore, learners in disadvantaged areas will continue to be deprived unless the Department of Basic Education (DBE) provides learners and teachers with suitable smartphones, tablets, or computers to access learning online [27]. Teachers who use ICTs often encounter weak, intermittent, or non-existent internet connections [28]. In such cases teachers are forced to revert to traditional methods of teaching [29]. In SA, regular loadshedding has become necessary due to aging electricity infrastructure and increased electrical energy demands [30-31]. In schools, power outages are especially disruptive to technology-enabled learning [32]. A successful eLearning project depends on the quality and ability of technical support. In some schools there is a lack of continuous and timely technical support provided by the technical department, which is often unreachable [33].

Restrictive Policies and Controls. It is the ICT policies of a country that determine what must be done to achieve its national goals in terms of technology adoption and use [34]. However, since the post-apartheid era, SA schools are seen as an autonomous body whereby governance and management decisions, such as policies, are decided by the stakeholders of the school essentially the Schools Governing Body (SGB) [35]. In a study conducted by Mwapwele et al. [36] on 24 rural schools across SA, it was found that 163 of the 197 teachers revealed that policy stated that their schools had prohibited students from using personal digital devices on school grounds. Furthermore, Mabhena [37] acknowledged free Wi-Fi was not embraced successfully at schools because of policies restricting phone use. Educators within the classroom struggle with the restrictive policies. On the one hand, there is a need to control the risky use of technology. However, on the other hand, learners need the freedom to utilize the technology to its full potential. Finding this balance is hard [38].

In summary, limited research can be found on school management views on internet adoption in schools. Most of the studies done by researchers have been interviews and surveys of teachers. Literature also tends to lean heavily on foreign countries' views regarding internet use for learners and the effects thereof. Taking that step up the chain of command is the objective of this study, by getting the views of the school management and in the process add literature from an SA context. This research will help other institutions understand the positive and negative outcomes of free Wi-Fi access for learners. This study will also be useful to education departments in other provinces looking to implement Wi-Fi in schools. Finally, this study also gives WCED a snapshot of what is going on in their schools, which could assist in proposing policy and planning changes.

3 Research Method

We present an interpretive multiple case study of five schools (S1-S5) performed in mid-2022. SA public schools are divided into five quintiles [39]. Underprivileged (poorest) schools are in Quintile 1, while economically privileged (richest) schools are in Quintile 5. We targeted WCED primary and high schools within these quintile levels, which received the rollout of Local Area Networks (LANs) consisting of Wi-Fi. We did case studies in Quintile 3 to 5 schools, the wealthier public schools. A summary of school characteristics is presented in Table 1. Prior to data collection, approval was obtained from the Commerce Ethics in Research Committee of the University of Cape Town and the WCED. Secondary data sources collected include various school policies, circulars and a briefing document which was provided by an IT Technician as shown in Table 2. Documents SSDA and SSDB were relevant to all cases.

Table 1. Case Descriptions.

Case ID	School type	Connectivity	Wi-Fi access (years)	School District	Teachers: Learners
S1	Quintile 3: No-fees Rural High School	Fair to Poor	5	Overberg	22:690
S2	Quintile 3: No-fees Urban Primary School	Good	6	North	25:695
S3	Quintile 4: No-fees Urban Primary School	Poor	4	Central	45:1080
S4	Quintile 5: Fee-paying Urban Primary School	Good	3	South	30:845
S5	Quintile 5: Fee-paying Urban High School	Below Average	7	North	34:820

Table 2. Secondary Sources.

Document ID	Description	Source	Document ID	Description	Source
S501	Phone Policy	Website	S503	Social media Policy	Email
S502	Cyber-safety Policy	Website	S504	IT Acceptable Use Policy	Email
SSDA	Circular Broadband	Website	S301	Code of conduct	Email
SSDB	School LAN Briefing Document	IT technician			
SSDA	Circular Broadband	Website			
SSDB	School LAN Briefing Document	IT technician	S302	Social media Policy	Email
			S401	Computer Room Policy	Email

For semi-structured interviews we targeted seven participants being either Principals, Deputy Principals or SGB Members. All seven participants signed consent forms and a guarantee of anonymity was provided to schools and participants. Their details and participant code indicating their school, are in Table 3. Interviews were either face-to-face at the school or via Microsoft Teams.

Table 3. Participants Interviewed

Partic-ipant	Role	Subjects and Grades taught	Years teaching	Years using WCG Wi-Fi
S1A	Head of Department	Accounting, Mathematics, CAT. Grade 11&12	12	5
S1B	Acting Deputy Principal	Mathematical Literacy. Grade 11 & 12	16	5
S2A	Teacher, SGB	Mathematics, English, Afrikaans, isiXhosa. Grade 1	28	6
S3A	Bursar, ICT Committee, ex SGB	None	6	4
S4A	Head of Department	Life Skills, English Afrikaans, Mathematics, isiXhosa. Grade 3	8	3
S5A	Principal	History. Grade 8	20+	7
S5B	Operations Manager, SGB	None	12	7

Audio recordings were uploaded to MS Word 365 for automated transcription and NVivo software was used for the analysis of the secondary sources and cleaned transcribed data. We conducted a hybrid inductive and deductive six-step thematic analysis approach [40]. The literature was used to develop the initial codebook. We iterated through the six steps. Step 1 was reading through the data. In step 2 the codes from the codebook were created in NVivo under the relevant research questions. In step 3 the text was coded to the relevant codes and new codes were added. In step 4 we created themes: as the researchers went through the data, adding codes to each theme, additional codes emerged from the data and were later elevated into themes. Step 5 - reviewing themes: a review of the themes was conducted at this stage by the researchers to verify the relevance of the data assigned to each theme. Step 6 - defining and naming themes: most themes were defined by the codebook, however new themes were defined and renamed as coding continued. Finally, presenting and discussing results: data was used by the researchers in this study to verify the findings.

The researchers attempted **member checks**, by having participants review and confirm that the data was captured correctly. Through interviewing and collecting documents data triangulation was attempted. The authors then performed **peer examination and debriefing**. These improve dependability and credibility of the analysis [41]. Yet the study had limitations with data collection. A case from each of the quintile schools was wanted, but schools were unwilling to participate in the research. In some cases, school principals wanted to delegate the interview to an Information Technology (IT) teacher who was not a member of management. Requests for school policies went unanswered and in two schools' request for member checks on transcribed data went unanswered. In S2 data triangulation was not done as only one data source was obtained.

4 Research Findings and Discussion

In this section we describe the constraints and outcomes of giving free Wi-Fi access to learners in schools from a school management perspective.

4.1 Infrastructure Constraints

Infrastructure constraints refer to constraints which affected the schools' resources such as access to electricity and ICT.

Inadequate Connectivity. Inadequate connectivity was viewed by all participants as one of the biggest constraints. Participant S4A states *"some teachers have problems with it (connectivity), but I don't."* This poor connectivity is confirmed by S1A who says *"our connectivity is poor, very poor. The Wi-Fi goes off a lot. The Wi-Fi signal is very poor."* Our findings are supported by prior literature where Mahlo [28] declares teachers who use ICTs often encounter weak, intermittent, or non-existent internet connections. Hence if teachers had problems connecting pupils would too.

Loadshedding. Loadshedding is another constraint, reducing access to learners. Five participants contributed to this theme. Participant S5B highlights the confusion within the school structures on the importance of have Wi-Fi that works. *"We were discussing load shedding and how it's damaging our devices and the chairperson of the SGB said you can teach without Wi-Fi and the deputy responded, no you can't"* (S5B). Additionally, (S4A) highlights that *"if loadshedding is during your class time then you forfeit your period for the week (no wi-fi connection), so learners don't get the makeup period for that hour that they have lost. Students would have to wait until the following week."* Literature supports our findings where loadshedding not only disables your Wi-Fi connection, but smartboards and desktop PCs too. Damage and maintenance of equipment are among the costliest aspects of loadshedding [30].

Insufficient Device Access. Insufficient devices reduce learners' access to Wi-Fi. Padayachee [26] mentions that some learners do not have access to devices such as tablets and smart phones which the participants also confirm. *"Another thing that we have discovered is that a lot of kids don't have cellphones, it's mommy's phone or daddy's phone"* (S1B). Similarly participant S3A adds that "*in terms of the computers, we only have one computer lab, they consist of 40 PCs*". This statement exposes the constraints schools are under when dealing with a lack of available PCs and needing to support a learner body in the several hundred.

Low Spec. Devices. A new finding that emerged from data was that of compatibility issues. Under certain instances, there is difficulty connecting to the Wi-Fi due to compatibility issues, as indicated by the following quotes, *"One uses her MacBook, and* *finds it extremely difficult to connect to the Wi-Fi"* (S2A). *"The low specifications on student devices struggles to connect to Wi-Fi and keep the connection"* (S5A).

Theft of School IT Resources. Another new theme to emerge from data was that of theft. Theft minimises learners' access to Wi-Fi. Furthermore, schools are also targeted. The Minister of Education of the Western Cape, noted that in the holiday period of December 2021 and January 2022, 34 schools reported incidents of burglary and vandalism. Items stolen included IT and audio- visual equipment, as well as electrical cables [3]. In terms of cable theft, when the cable is cut it brings Wi-Fi access and learning to a grinding halt, affecting either a section of the school or the whole school itself, as the following quotes show. *"There was a time when we were off through the theft that damaged a few cables"* (S2A). Similarly, S3A adds that *"Our fibre has been on and off because our fibre cables are being damaged by stealing."*

4.2 Organizational Constraints

Organizational constraints refer to constraints which prevent or reduce use of the internet and are now discussed. The literature has instances of inadequate technical support for teachers and learners [33]. In this study the support was not seen as a constraint as the WCED service desk was seen as available and proactive. The constraints noted are now discussed.

Restrictive Policies. Restrictive policies are one of the biggest constraints to learners' accessing the Wi-Fi. As schools lack devices, phones are a device which could allow learners to access the internet. Free Wi-Fi use at schools has been reported as low because of policies restricting phone use [37]. Our findings confirm that as all participants mentioned school policies not allowing phone use in classrooms. S4A states that *"the students have to notify their teachers that they have a cellphone, then the teacher keeps it in the box in front of the class."* S5A adds that *"we take the cellphones in and then at the end of the day, we hand the cellphones back to the students."*

Insufficient Teacher Competency. Teacher competency is a constraint impacting learners' access to Wi-Fi. Schools are attempting to implement more technology-based lessons, but teachers must first acquire the necessary skills, which can be obtained through training. Successful ICT use in education is influenced by teaching attitude, teacher skill, and inexperience with ICT [23]. While some schools offered training, not all did, as evidenced by the following quotes. *"We had training, all the teachers had training when the Wi-Fi was setup"* (S2A). Participant S3A stresses that *"there was definitely no training. We just received the booklet on how to connect through different devices. Most teachers were lost"* (S3A).

Lack of Awareness of Support. A new theme to emerge was the school management's lack of awareness of learner accounts and support, they were unaware that learner login credentials could be accessed through the school management system. *"But we are not familiar when a learner, if he's got any issues with the login, if they can log a call with service desk to assist and help"* (S5A). *"So, I mean, if I knew that they actually could have their own password, I would have done that administrative duty to the department and get them their passwords"* (S3A). Secondary sources on the internet noted that the WCED had sent out a circular to schools (SSDA) with a link to their WCED ePortal where learner logins, can be found.

Lack of Device Control. Educators want to control and manage unrestricted, risky use of technology while utilizing its potential at the same time [38]. In our study the lack of device control was identified as the primary constraint preventing learners' access to Wi-Fi. Schools would like to monitor what learners access and, if necessary, restrict access to sites and applications. Participant S3A *"feels that if schools we're going to allow the learners to connect to the Wi-Fi, it just needs to be in a controlled environment."* Similarly, participant S1A adds that *"there are so many people on the Wi-Fi updating and you can't control it."* It appears that schools lack the ability to restrict which Wi-Fi connections are allowed and which are not.

4.3 Positive outcomes of Wi-Fi to learners

Positive outcomes refer to benefits of the Wi-Fi for learners and are now discussed.

Access to Online Information. The internet is the easiest and quickest way of finding relevant information that meets users' needs [8]. Hence access to Wi-Fi opens doors to information that learners from impoverished areas have not had much experience with. Access to information is critical when conducting research for projects and assignments. Furthermore, if a learner is unclear or does not understand a specific topic, they can access that information online. This was confirmed by our participants with S1B saying that *"learners don't have access at home, so Wi-Fi allows them access to many online resources."* Additionally, *"in the class, if there's some term you don't understand or you need more information on, using your phone with Wi-Fi access you can get the information you are looking for"* (S5A).

Enhanced Communication and Collaboration. Learners could communicate and collaborate through online communication and collaboration applications, which allows them to stay motivated and keep their focus on quality work [12]. All participants mentioned that the Wi-Fi is commonly used for social media access such as WhatsApp, TikTok, Facebook, and Instagram. However, in our study the online communication by learners was mostly geared towards social communication rather than educational communication. *"I need my learners to have a cellphone at school, because we help them create a Gmail address for communication."* (S1A). Additionally participant S5B says that *"learners will also utilize the school Wi-Fi to access social media. After school* *when they are waiting for their parents to pick them up, they usually access social media to keep busy"* (S5B).

Accessing Online Educational Content. Enabling learner access to educational websites such as Khan Academy provides access to subject-specific content [10]. All participants noted that access to educational websites and content, for example the WCED ePortal is a benefit. *"For learners having Wi-Fi, there's more learning material they can access"* (S5B). *"The Department of Education has the ePortal which is free to the parents and learners to access any textbooks, reading material, they even have an online library"* (S4A).

Accessing Game-based Learning Within Classrooms. Game-based learning has academic benefits [11]. According to the participants, Wi-Fi enables access to online games which teachers use for learning such as Greenshoots (for mathematics), Kahoot and games on Youtube. This is seen to make lessons more fun and engaging. This theme was contributed to by five participants. *"There are some teachers that's making use of Wi-Fi, like they want to play Kahoot"* (S5A). *"We make use of YouTube interactive games online for the children to participate in the lessons"* (S4A). *"I know the Grade 3 to 7, they do the Greenshoots now in the lab"* (S2A).

To Control Learners and Reduce Learner Stress. A new theme to emerge which two participants highlighted was that online content can assist with discipline issues and stress levels. *"Even entertainment, breaktime, I want to listen to music, we know it calms you down. Yes, it will help with discipline"* (S5A). *"Besides improving the reading skills, it could also help them (learners) to relax"* (S2A).

4.4 Negative outcomes of Wi-Fi to learners

Negative outcomes, which we now describe, refer to drawbacks of Wi-Fi for learners.

Cyberbullying. Cyberbullying is viewed as a negative outcome of learners having access to Wi-Fi. Learners can insult and inflict pain on one another through cyberbullying [18]. This outcome was confirmed in that four of the five cases reported cyberbullying incidents at their school. Participant S1B stated that *"cyberbullying is one of our biggest challenges this year and especially in our grade 8 and 9's. Girls make use of cyberbullying especially."* Similarly participant S5A mentions that they are *"currently working with the Department of Education or the Metro North, to run a project in the school regarding awareness, regarding cyberbullying, because it is a continuous problem."*

Internet Abuse and Accessing Pornography. Internet abuse is accessing the internet beyond what the acceptable use policies authorize [17]. Literature notes that students abuse school internet by using it for social media [16] Five participants mentioned this as a concern. Students can also access inappropriate content such as pornography which is seen as a major concern in all schools. Currently, there are no system restrictions and *"most learners are on WhatsApp, YouTube, and TikToK. It is free reign for everyone"* (S3A). *"I don't blame the principal for not allowing them to bring cell phones ... you can't control what they search at the end of the day, and we have had instances where they (learners) wanted to download porn games"* (S3A). *"You know young kids, especially the boys accessing naughty sites"* (S1B).

Being Distracted from Learning. Wi-Fi access can cause distractions in the classroom [38], learners can also become addicted to internet use, which can negatively impact learning [8]. Although participants noted excessive social media use, addiction was not mentioned by a single participant. This could be due to the limited access given to learners. Yet distraction was regarded as a negative outcome of learners having access to Wi-Fi with five participants noting this concern. *"Some learners go on Facebook and others video call during lessons"* (S1A). *"So, the other one is of course their attention span, of the learners in the class. In other words, if they've got a device with them, it will take away that needed attention from the lesson that you need in class for that specific subject"* (S5A).

4.5 Summary of Findings

The purpose of our study was to describe the constraints and outcomes of giving free Wi-Fi access to learners in Western Cape schools from a school management perspective. Fig. 1 summarises those constraints and outcomes identified by the school management linked to giving free Wi-Fi access to learners in Western Cape schools.

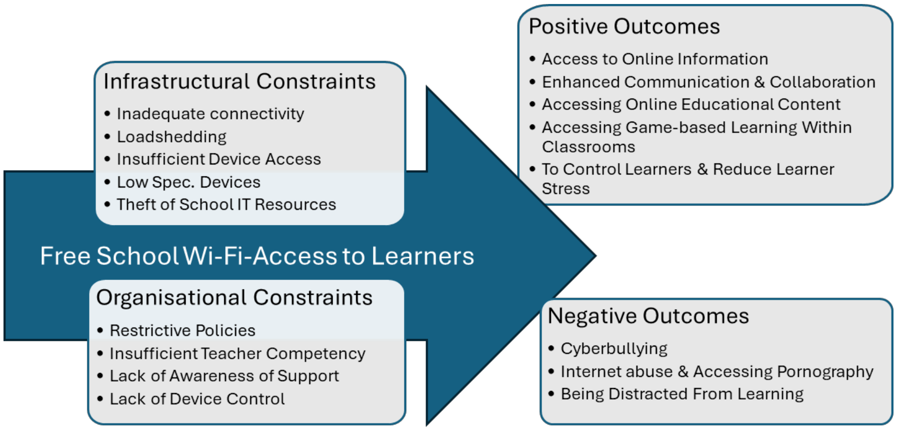

Infrastructural Constraints
• Inadequate connectivity
• Loadshedding
• Insufficient Device Access
• Low Spec. Devices
• Theft of School IT Resources
Free School Wi-Fi-Access to Learners
Organisational Constraints
• Restrictive Policies
• Insufficient Teacher Competency
• Lack of Awareness of Support
• Lack of Device Control
Positive Outcomes
• Access to Online Information
• Enhanced Communication & Collaboration
• Accessing Online Educational Content
• Accessing Game-based Learning Within Classrooms
• To Control Learners & Reduce Learner Stress
Negative Outcomes
• Cyberbullying
• Internet abuse & Accessing Pornography
• Being Distracted From Learning

Fig. 1. Constraints and outcomes linked to giving free Wi-Fi access to learners by their schools.

5 Conclusion

In this paper we described the constraints and outcomes of giving free Wi-Fi access to learners in Western Cape schools from a school management perspective. The Western Cape Government started rolling out free Wi-Fi to schools in 2015 [2], however literature found that mostly teachers were making use of the Wi-Fi service and that learners' access were limited. We found that three of the five schools do not give Wi-Fi access to learners. The study also highlighted various constraints impacting the reasoning for learners currently not using the Free Wi-Fi. Infrastructural constraints reported in prior literature include inadequate connectivity, loadshedding and insufficient device access, while new themes included low spec. devices and theft of IT resources. The dominant constraint identified was restrictive school policies and lack of device control. Other organizational constraints include insufficient teacher competency, and a new theme was the lack of awareness of support offered learners.

Furthermore, our study also highlighted both positive and negative outcomes for learners using the school's Wi-Fi. Game-based learning within classrooms and access to online educational content are some positive outcomes. An interesting new theme was the use of the internet to reduce learner stress. Yet management are concerned about the negative outcomes with cyberbullying, internet abuse, pornography and being distracted from learning all being reported as real concerns. It seems these negative outcomes are outweighing the benefits, and that school management will allow access to free Wi-Fi, but only when it is a controlled environment.

The study will be useful to education organizations and government decision makers in understanding the position schools are in and what schools require to give Wi-Fi access to learners. This study will also be useful to education departments in other provinces looking to implement Wi-Fi in schools. Finally, the paper identifies four key areas where investment is needed for the WCED, namely tools to manage personal devices, better awareness of available support, increased bandwidth, and adequate teacher training.

The paper has limitations in that schools are mostly from lower-income areas within their suburb. None of the schools which participated were affluent schools, where every learner has a device with a strong internet connection. Further analysis could have indicated if those schools had similar challenges in terms of managing phones and what tools they employ. This study also only had one school from the rural areas. Future research should be conducted with a broader range of schools, including private schools.

Acknowledgments. Funding support to write this study is acknowledged from the National Institute for the Humanities and Social Sciences. This research was conducted as part of a BCom Honours in Information Systems empirical research.

References

1. Salahuddin, M., Gow, J.: The effects of Internet usage, financial development and trade openness on economic growth in South Africa: A time series analysis. Telematics and Informatics. 33(4), 1141-1154 (2016). https://doi.org/10.1016/j.tele.2015.11.006.
2. WCED.: WCED announces details on e-learning “smart schools” project, https://www.westerncape.gov.za/news/media-release-wced-announces-details-e-learning-%E2%80%9Csmart-schools%E2%80%9D-project.
3. WCED.: 2022 Western Cape Education Budget Speech, https://wcedonline.westerncape.gov.za/news/2022-western-cape-education-budget-speech.
4. Katsidzira, K., Seymour, L.F.: Factors Impacting Using the Internet for Learning: The Digital Divide in South African Secondary Schools. In: Wells, G., Nxozi, M., Tait, B. (Eds.), ICT Education. SACLA 2020. Communications in Computer and Information Science. 1518. pp. 1-18. Springer, Cham. (2020). https://doi.org/10.1007/978-3-030-92858-2_1.
5. Pahlavan, K., Krishnamurthy, P.: Evolution and Impact of Wi-Fi Technology and Applications: A Historical Perspective. International Journal of Wireless Information Networks. 28(1), 3–19 (2021). https://doi.org/10.1007/s10776-020-00501-8.
6. Yang, M., Li, B., & Yan, Z.: MAC Technology of IEEE 802.11ax: Progress and Tutorial. Mobile Networks and Applications. 26(3), 1122–1136 (2021). https://doi.org/10.1007/s11036-020-01622-3.
7. Chigona, A.: Digital fluency: necessary competence for teaching and learning in connected classrooms. The African Journal of Information Systems. 10(4), 366-379 (2018).
8. Puspita, R., Rohedi, D.: The impact of internet use for learners. In: Proceeding of the IOP Conference Series: Materials Science and Engineering 25th-26th October, pp. 1-7 (2018). https://doi:10.1088/1757- 899X/306/1/012106.
9. Kantharia, M.: Online School Education in India during Coronavirus Pandemic: Benefits and Challenges. Research Journal of Humanities and Social Sciences. 11(2), 99-103 (2020).
10. Malamud, O., Cueto, S., Cristia, J., Beuermann, D.: Do children benefit from internet access? Experimental evidence from Peru. Journal of Development Economics, 138, 41–56 (2019). https://doi.org/10.1016/j.jdeveco.2018.11.005.
11. Foster, A., Shah, M.: Principles for Advancing Game-Based Learning in Teacher Education. Journal of Digital Learning in Teacher Education. 36(2), 84–95 (2020). https://doi.org/10.1080/21532974.2019.1695553.
12. Maican, C., Cazan, A., Lixandroiu, R., Dovleac, L., Maican, M.: Exploring the factors influencing the use of communication and collaboration applications. Journal of Organizational and End User Computing (JOEUC). 33(4), 94-124 (2021). DOI:10.4018/JOEUC.20210701.oa5.
13. Yebowaah, F.: Internet Use and its Effect on Senior High School Students in Wa Municipality of Ghana. Library Philosophy & Practice. 1-30 (2018).
14. Kayalar, F.: Cross-Cultural Comparison of Teachers' Views upon Integration and Use of Technology in Classroom. Turkish Online Journal of Educational Technology-TOJET. 15(2), 11-19 (2016).
15. Szymkowiak, A., Melović, B., Dabić, M., Jeganathan, K., Kundi, G.: Information technology and Gen Z: The role of teachers, the internet, and technology in the education of young people. Technology in Society. 65, 101565 (2021). https://doi.org/10.1016/j.techsoc.2021.101565.
16. Almasi, M., Machumu, H., Zhu, C.: Internet use among secondary schools learners and its effects on their learning. In: Proceedings of INTED2017 Conference 6th – 8th March, pp. 1-12 (2017). https://DOI:10.21125/inted.2017.0680.

17. Ezema, M., Inyama, H.: An assessment of Internet Abuse in Nigeria. West African Journal of Industrial and Academic Research. 4(1), 168-172 (2012).
18. Scholtz, B., Van Turha, T., Johnston, K.: Internet visibility and cyberbullying: a survey of Cape Town high school learners: informatics for development. The African Journal of Information and Communication. (15), 93-104 (2020).
19. Claro, M., Nussbaum, M., López, X., Contardo, V.: Differences in views of school principals and teachers regarding technology integration. Journal of Educational Technology & Society. 20(3), 42-53 (2017).
20. Mogas, J., Palau, R., Fuentes, M., Cebrián, G.: Smart schools on the way: How school principals from Catalonia approach the future of education within the fourth industrial revolution. Learning Environments Research. 25, 875–893 (2022). https://doi.org/10.1007/s10984-021-09398-3.
21. Goh, E., Sigala, M.: Integrating Information & Communication Technologies (ICT) into classroom instruction: teaching tips for hospitality educators from a diffusion of innovation approach. Journal of Teaching in Travel & Tourism. 20(2), 156-165 (2020). https://DOI:10.1080/15313220.2020.1740636.
22. Dias, L., Victor, A.: Teaching and Learning with Mobile Devices in the 21st Century Digital World: Benefits and Challenges. European Journal of Multidisciplinary Studies. 2(5), 26-34 (2017). https://doi.org/10.26417/ejms.v5i1.p339-344.
23. Ndlovu, N., Moll, I.: Teachers, Technology and Types of Media: Teaching with ICTs in South Africa. African Educational Research Journal. 4(3), 124-130 (2016).
24. Winter, E., Costello, A., O'Brien, M., Hickey, G.: Teachers' use of technology and the impact of Covid-19. Irish Educational Studies, 40(2), 235-246 (2020). https://doi.org/10.1080/03323315.2021.1916559.
25. Alexander, H.: Investigating technology acceptance in public secondary schools in Pretoria. Doctoral dissertation, University of Pretoria (2019).
26. Padayachee, K.: A snapshot survey of ICT integration in South African schools. South African Computer Journal, 29(2), 36-65 (2017).
27. Mkhlze, T., Davids, M.: Towards a digital resource mobilisation approach for digital inclusion during COVID-19 and beyond: A case of a township school in South Africa. Educational Research for Social Change. 10(2), 18-32 (2021). http://dx.doi.org/10.17159/2221-4070/2021/v10i2a2.
28. Mahlo, L.: The capabilities necessary for effective ICT integration in teaching at two public primary schools in Khayelitsha in the Western Cape. Doctoral dissertation, Cape Peninsula University of Technology (2020).
29. Saal, P., Graham, M., Van Ryneveld, L.: Integrating educational technology in mathematics education in economically disadvantaged areas in South Africa. Computers in the Schools. 37(4), 253-268 (2020). https://doi.org/10.1080/07380569.2020.1830254.
30. Goldberg, A.: The economic impact of load shedding: The case of South African retailers. Doctoral dissertation, University of Pretoria (2015).
31. Mabunda, N.: Use of Photovoltaic Energy to Minimize the Impact of Load-shedding in South Africa. In: International Conference on Electrical, Computer, and Energy Technologies (ICECET), pp. 1-4 (2021). https://doi.org/10.1109/ICECET52533.2021.9698720.
32. McCain, N.: Education sector concerned as impact of load shedding cuts down valuable teaching, studying time (2022). https://www.news24.com/news24/southafrica/news/education-sector-concerned-as-impact-of-load-shedding-cuts-down-valuable-teaching-studying-time-20220920.

33. Nawaz, A., Khan, M.: Issues of technical support for e-learning systems in higher education institutions. International Journal of Modern Education and Computer Science. 4(2), 38-44 (2012). https://doi.org/10.5815/ijmecs.2012.02.06.
34. Gibson, D., Broadley, T., Downie, J., Wallet, P.: Evolving learning paradigms: Re-setting baselines and collection methods of information and communication technology in education statistics. Journal of Educational Technology & Society. 21(2), 62-73 (2018).
35. Williams, C.: Distributed leadership in South African schools: possibilities and constraints. South African Journal of Education, 31(2), 1-11 (2011). DOI:10.15700/saje.v31n2a421.
36. Mwapwele, S., Marais, M., Dlamini, S., Van Biljon, J.: Teachers' ICT adoption in South African rural schools: a study of technology readiness and implications for the South Africa connect broadband policy. The African Journal of Information and Communication. 24, 1-21 (2019). https://doi.org/10.23962/10539/28658.
37. Mabhena, Z.: Domestication of free Wi-Fi amongst high school learners in disadvantaged communities in the Western Cape, South Africa. Master's thesis, University of Cape Town (2017).
38. Morris, P., Sarapin. S.: Mobile phones in the classroom: Policies and potential pedagogy. Journal of Media Literacy Education. 12(1), 57-69 (2020). https://doi.org/10.23860/JMLE-2020-12-1-5.
39. Maistry, S., Africa, I.: Neoliberal stratification: The confounding effect of the school poverty quintile ranking system in South Africa. South African Journal of Education, 40(4), 1–9 (2020). https://doi.org/10.15700/saje.v40n4a1872.
40. Labra, O., Castro, C., Wright, R., Chamblas, I.: Thematic analysis in social work: A case study. Global Social Work-Cutting Edge Issues and Critical Reflections. 1-20 (2020). https://doi.org/10.5772/intechopen.89464.
41. Anney, V.: Ensuring the quality of the findings of qualitative research: Looking at trustworthiness criteria. Journal of emerging trends in educational research and policy studies. 5(2), 272-281 (2014).

Stakeholder Theory and Enterprise Architecture in Serious Games: An Integrative Review

Lance Bunt[1][0000-0003-0455-8493] Estelle Taylor[2][0000-0003-2848-7829] Japie Greeff[3][0000-0002-5743-8922]

[1, 2] North-West University, Unit for Data Science and Computing (UDSC), South Africa
[3] North-West University, Optentia, South Africa
Lance.Bunt@nwu.ac.za

Abstract. This study addresses the oversight of stakeholder-focused strategies in traditional serious game frameworks by proposing a stakeholder-centred approach to serious game design. Through a comprehensive review and synthesis of literature across serious games, stakeholder theory, and enterprise architecture, this paper highlights the importance of integrating stakeholder theory and enterprise architecture principles in the development of serious games. An integrative literature review process—screening, extracting, synthesizing, and evaluating data—led to the development of a conceptual framework for serious game design. This framework is characterized by its flexibility, stakeholder-centeredness, goal orientation, and supportive nature, designed to overcome common challenges in serious game development. It prioritizes the needs and interests of all stakeholders, including players, designers, investors, and regulators, emphasizing the application of stakeholder theory and enterprise architecture methods to enhance serious game development. The proposed framework facilitates the prioritization of stakeholder needs and alignment with organizational goals, thereby improving player experiences and ensuring scalability and security.

Keywords: Integrative Literature Review, Serious Games, Stakeholder Theory, Enterprise Architecture, Conceptual Framework.

1 Introduction

1.1 Background

The process of developing serious games (SGs) is not all fun and games [1-4]. It is costly, time-consuming, and laborious to create successful artefacts that both accompany and edify learning or do more than solely entertain [5-8]. SGs, moreover, have a specific educational, training, health, or social determination [9]. Stakeholders in SGs—those invested in or impacted by the project's success—face challenges, including communication barriers that can be mitigated with effective planning [10]. The production of SGs relies on collaborative effort across various domains, demanding various skills and roles [11]. Effective communication and stakeholder management, supported by an integrated framework, are crucial for achieving project goals and ensuring stakeholder productivity and satisfaction before, during and after a project's lifecycle [12].

1.2 Problem statement & research gap

Serious games (SGs) are recognised for their significant impact on education, training, and healthcare by enhancing learning outcomes, improving technical skills, and providing interactive and engaging experiences across various domains. The development of SGs for educational, training, health, or social purposes involves multiple stakeholders, each bringing distinct competencies, roles, responsibilities, and objectives to the table [13,14,15]. A prevalent issue in this multidisciplinary field is the lack of effective communication strategies and adequately structured stakeholder engagement [16,17]. This deficiency can lead to inefficiencies and dissatisfaction, potentially compromising project success. A significant gap identified in the literature is the absence of a comprehensive framework that integrates Stakeholder Theory (ST) [18,19] and Enterprise Architecture (EA) principles [20] in the SG design and development process. Current SG development frameworks overlook the stakeholder perspective, evidenced by:

- Insufficient mechanisms to identify, prioritise, and engage SG stakeholders, risking the neglect of crucial needs and interests, which can lead to conflicts and project failure [18,19].
- A lack of clear role and responsibility definitions for stakeholders in SG development, leading to confusion and miscommunication [20].
- An absent value proposition to secure the ongoing commitment and engagement of SG stakeholders [21,22].

This paper systematically categorizes current research in SGs ①, ST ②, and EA ③, identifying key sources and significant studies. The analysis reveals methods for enhanced stakeholder management within SG teams ②, elucidates various factors influencing SG development processes ①, and employs EA literature to bridge these domains, offering constructive rationale and practices ③. The Open Group Architecture Framework (TOGAF) is also explored in this context. The outcome of this integrative review defines the requirements for a stakeholder-centric SG development framework, which is further discussed in Section 7.

The central research question addressed in this study is: *How can the domains of serious games (SGs), stakeholder theory (ST), and TOGAF enterprise architecture (EA) be integrated to provide a stakeholder-centred conceptual framework for the design of SGs?*

2 Methods

The integrative literature review (IRL) research strategy is pursued because it synthesises and evaluates existing research studies. The method, moreover, helps develop understanding for the conceptual SG design framework and informs evidence-based practice going forward.

2.1 Integrative literature review strategy

An ILR is a type of study that evaluates, analyses, and synthesises representative literature on various topics in a unified manner so as to produce new frameworks and perspectives on the issues explored [24]. Three components make up the integrative literature review in this research: a typology analysis, theory mapping, and instruction writing. Firstly, the review classifies the subjects of ST, SG development and EA by way of a typology put forward by Onwuegbuzie and Frels [25]. Secondly, the review captures the essence of the subjects by way of theory maps [26]. Thirdly, a (a) timeline for each pillar is drawn, (b) both meta- and sub-concepts are unified, and (c) the relationships between the domains are distinguished, as suggested by Torraco [24]. This paper presents the results of these processes.

Literature review typology. Four types of integrative literature reviews were developed by Onwuegbuzie and Frels [25]: theoretical, empirical, methodological, and mixed methods. The purpose of this theoretical literature review is to integrate and synthesise the domains of ST, SGs, and EA.

Theory maps. Repko and Szostak's [26] theory maps offer a representation of the connections between theories, concepts, and phenomena. The following elements comprise the theory map for the conceptual framework for the design of SGs: Firstly, serious games (SGs) ① are designed for non-entertaining purposes and are utilized by organizations to provide a more engaging and interactive learning environment. Secondly, stakeholder theory (ST) ② emphasizes that organizations should consider the interests and expectations of stakeholders, including employees, customers, suppliers, shareholders, and broader society, during the decision-making process. Thirdly, enterprise architecture (EA) ③, through frameworks like TOGAF, offers a comprehensive methodology for designing, planning, implementing, and managing the information technology architecture of an organization. The stakeholder-centred design methodology integrates these principles by ensuring that the design of SGs ① takes into account the diverse interests and expectations of all stakeholders involved in their development and implementation.

Integrated Literature Review Guidelines. The steps outlined in Torraco's [23] guidelines for conducting ILRs are as follows:

To develop a stakeholder-centred conceptual framework for the design of SGs, we first formulated the research question: How can the domains of SGs, stakeholder theory, and TOGAF enterprise architecture be integrated to provide such a framework? We then conducted a thorough search of relevant literature using various databases, including Scopus, Web of Science, and Google Scholar, with keywords such as "serious games," "stakeholder theory," "TOGAF," and "enterprise architecture." The chosen studies were assessed based on their quality and relevance to the research question, including only peer-reviewed articles published in reputable journals. Data from these selected studies were extracted and analysed to determine common themes and patterns. Finally, the findings were synthesized to develop the requirements and characteristics of a stakeholder-centred conceptual framework for the design of SGs.

The ILR method, in the context of this research, involves synthesising and analysing multiple sources of literature to develop a thorough understanding of a research topic.

3 Integrated literature review

The typology put forward by Onwuegbuzie and Frels [25] pronounces the justification scholars and researchers utilise when undertaking integrated literature reviews. The first step in this rationale involves informing the topic.

3.1 Informing the pillars of study

Table 1 presents the first step in the integrative literature review process, informing the three pillars under investigation:

Table 1. Integrative literature review overview.

Query	Detail & Explanation		
What is the origin of the topics explored?	The provenance of the three (3) domains or pillars are as follows: • **Serious Games** ①**:** C.C. Abt executed seminal work and coined the term in 1970. B. Sawyer popularised it in 2002; • **Stakeholder Theory** ②**:** Finds its roots in work done by E.R. Freeman in the influential work "Strategic management: A stakeholder approach" in 1984. The field and approach has seen application i mains since; and • **Enterprise Architecture** ③**:** Born in the 1960s from architectural manuscripts on Business Systems Planning (BSP), initiated by IBM and P. Duane Walker.		
Are the topics clearly defined?	**Serious games** ① are applications of gaming technology, process and design that emphasise problem solving, focus on the elements of learning, incorporate assumptions necessary for workable simulations, reflect natural and non-perfect communication, and go beyond pure entertainment [30].	**Stakeholder Theory** ② provides business perspectives regarding how an organisation should generate and maintain value for their customers, communities, shareholders, and other interest groups; in various disciplines and use-cases, including corporate strategy, finance, management, business ethics and marketing [29].	**Enterprise Architecture** ③ is a discipline that describes and manages organisational structure, procedures, systems, and technology in an integrated way. Moreover, it investigates the effects of organisational change to design models for enhanced stakeholder analysis and reporting [31].
Is the scope of the review known?	Literature in the SG ① arena are vetted and limited by their focus/foci. For example, research that focuses on [serious] game development, planning, design, conceptualisation, delivery, methodology, framework(s), theory, workflows, teams and processes will be considered for review. Moreover, case studies and practical guidelines for SG ① development will be prioritised in this corpus. Literature, here, concerns SG ① creation; not application.	Literature in the ST ② domain are selected according to its feasibility and purpose. ST ② research pertaining to and affecting stakeholder identification, selection, analysis, prioritisation, management, communication, and optimisation make up the selected literature for this theoretical pillar. Application, then, is the central aspect of this research pillar. This literature coupled with EA ③ inform framework function.	Literature in the EA ③ sphere are chosen based on relevance for the research. Purposeful attention will be given to TOGAF, Control Objectives for Information and Related Technology (COBIT) and Information Technology Infrastructure Library (ITIL) in this study; as they are significant to the research context (serious game development) and key concepts (stakeholder interaction, management, tooling, and optimisation) under investigation. Literature in the EA ③ corpus is therefore limited to guidelines from these three standards/methods.
Which chosen stance(s) are applicable?	A framework that is diagnostic, flexible, informative, repeatable, and sustainable (among other characteristics) would abet and support stakeholders who actively develop SGs ①. An applied framework aimed at informing best practice is anticipated to meet this need. Determining factors for best practice is the basis for this practice.	SG stakeholders generally operate in an ad-hoc, spontaneous, and impromptu manner when developing artefacts and media. ST ② can mitigate this operational oversight through various pertinent and functional theories, approaches and techniques for stakeholder analysis, engagement, and management.	The selected enterprise architecture frameworks should inform and support both short and long term tactical and operation SG development work. This involves setting up guidelines for intelligent system planning, implementation, and maintenance. EA ③ concepts are vital to operationalising the applied framework in the field.
How is literature selected?	Theoretical foundation → Generally related studies → Similar studies (topics, populations, research designs) → Research Gap		
Which keywords are used?	Serious game development, serious game planning, serious game design, serious game conceptualisation, serious game delivery, serious game methodology, serious game framework(s), serious game theory, serious game workflow(s), serious game team(s), and serious game management.	Stakeholder identification, selection, analysis, prioritisation, management, communication, effectiveness, planning, stakeholder theory, stakeholder approaches and optimisation.	Enterprise Architecture, TOGAF, COBIT, ITIL, Agile, Application Architecture, Architecture Framework, Assessment Metric, Business Capability (Modelling), Enterprise principals, Key Performance Indicator(s), Lifecycle, Scrum, Standard(s), Supply Chain Management (SCM), View, Viewpoint.
How is literature reviewed?	Reading only titles → Reading only abstracts → A staged review (certain sections/topics) → Complete reading of relevant literature source (staged review) → Analysis of literature integrated into paper		

Query	Detail & Explanation
Databases searched?	Web of Science, SCOPUS, JSTOR and Google Scholar
Inclusion Criteria?	The literature review should cover at least one of the three key theoretical domains: Serious Games, Stakeholder Theory, or Enterprise Architecture. It should include a mix of empirical and theoretical contributions, such as peer-reviewed articles, book chapters, conference proceedings, and white papers, provided they are in English or accompanied by an English translation. Additionally, the sources should be recent, ideally published within the last ten years, to reflect current practices and theories. The methodological rigor of empirical studies, including result validity, sample size, and statistical methods, must also be assessed.

Table 2 provides more information on each pillar, including modes, applications, objectives, methodologies, stakeholders, and landmark studies:

Table 2. Integrative **literature** review: Informing the pillars of study.

Pillar	Details
Serious games	Modes: Analogue/Digital Applications: Learning, Therapy, Social Control, Research Objectives: Teach skills, Disseminate information, Instil attitudes Methodologies: Learning methods, Information structures, Game features, Evaluation Stakeholders: Diverse positions, Activities, Specializations Landmark studies: • Abt (1970) • Hill et al. (2006) • Susi et al. (2007) • Aldrich (2009) • Breuer & Bente (2010) • Deterding et al. (2011) • Laamarti et al. (2014)
Stakeholder theory	Focus: Economic Value Creation Collaboration: Mutual, Deliberate, Voluntary Participants : Corporations and Individuals Value Production: For all stakeholders, not just shareholders Landmark studies: • Freeman (1984) • Goodpaster (1991) • Donaldson & Preston (1995) • Mitchell et al. (1997) • Phillips (2003) • Palmer (2015)
Enterprise architecture	Development Phases: Business Systems Planning, Early EA, Modern EA Problems Addressed: (i) System Complexity (ii) Suboptimal Business Alignment Core Foci of EA: Technology, Socio-technology, Ecotechnology Key Frameworks:TOGAF, COBIT, ITIL Landmark studies: • BSP (1975) • Spewak & Hill (1992) • Zachman (1997) • The Open Group (2011) • ISACA (2018) • AXELOS (2019)

3.2 Relationship between theoretical pillars

Upon initiating the ILR, it is important to note the views regarding the relationships between the three theoretical pillars under investigation. The distinction to take note of resides in the degree of interaction and integration among the three pillars. The rigid (pre-ILR) view sees them as distinct components with defined roles, whereas the integrated (post-ILR) view recognises them as interconnected elements that continuously interact and influence one another, resulting in a stakeholder-focused, technically robust, and adaptable SG development process.

The pre-ILR view sees the pillars in the following way:

- SGs concentrate solely on the design and gameplay elements to accomplish the desired goals, such as education, training, or social motivation [32].
- Stakeholder Theory focuses on the identification, prioritisation, and management of stakeholders, ensuring that their needs and interests are met without directly influencing the design or development of SGs [33].
- Enterprise Architecture is the structure that organises and aligns the technological infrastructure of the game with organisational strategies and processes. However, it is rigid and does not adapt to the changing needs of game development [34].

The post-ILR view regards the pillars as interconnected theories that dynamically influence one another throughout the SG development process:

- Integration with SGs: The development of the game is a dynamic system that evolves in response to the input of stakeholders and within the framework of the established architecture. As the game evolves, Stakeholder Theory and EA principles actively shape the development process [35].
- Stakeholder involvement and participation: Stakeholders are active participants throughout the entire life cycle of a game, from conception to design, development, testing, and deployment. Their ongoing participation ensures the game's continued relevance and adaptability to shifting needs and contexts [36].
- Interaction with Enterprise Architecture: The architecture is flexible and adaptable to accommodate changes based on feedback from stakeholders and evolving development requirements. It adapts to the game, ensuring that it remains technically sound and efficient even as requirements and game objectives change [37].

3.3 Narrowing the pillars of study

Narrowing SGs. SGs ① have been implemented in numerous settings, including healthcare, education, military training, and public policy [38]. Multiple benefits of using SGs have been demonstrated, including increased engagement, motivation, and knowledge retention [39]. However, designing and implementing SGs can be difficult, particularly when it comes to balancing the game mechanics with the learning objectives [28]. SGs ① are games that serve a purpose other than pure entertainment [40]. SGs ①, moreover, have been used in healthcare to simulate medical procedures, educate patients about disease management, and train healthcare professionals [41]. Various subjects, including mathematics, science, and languages, have been taught using

SGs in education [42]. SGs ① have been used in corporate training to improve employees' skills, knowledge, and attitudes [43].

Designation of serious game chosen: Edifying artefacts, tools and game-based initiatives crafted by development teams which incorporate ludic activity for a specific purpose, format, genre, interaction style and application area.

Current literature in the field of SGs ① remains inconclusive with regards to provisions for effective SG ① design procedures. According to Lameras et al. [44], this is due to various complications: (a) Aligning the roles of stakeholders (developers and teachers) is a constant challenge; (b) Dialogue between stakeholders is not always assured or dependable; and (c) The key role of practitioners or "content experts" (such as teachers) in ensuring learning activities, feedback and outcomes are carried out in the final SG are convoluted by "iterative and participatory methods frequently adopted". Endeavouring to tackle these issues, Lameras et al. [44] recommend that SGs be developed with learning schemes and subject matter in mind to enrich the learning experience of players.

Narrowing stakeholder theory. Stakeholder theory is a body of theory relating to business ethics and management methods which emphasise existing interactions between an organisation and its workers, customers, suppliers, investors, communities, and other stakeholders. This approach advocates for the provision of value for every stakeholder. ST ② emphasises the significance of considering all stakeholders' needs and interests when making decisions [29]. ST is especially applicable to the context of SGs, as they are frequently designed with a specific audience in mind, and it is essential to take their needs and interests into account.

A stakeholder is any individual, group or entity recognised by their potential to affect or be affected by business functioning and the realisation of organisational objectives.

ST ② proposes that organisations should consider the interests and expectations of diverse stakeholders in their decision-making [29]. An organisation's stakeholders include employees, customers, suppliers, shareholders, and society [29]. The design of SGs should consider the interests and expectations of all parties involved in the development and implementation of the game. For example, the design of a SG for corporate training should consider the needs of employees, trainers, and the organisation's leadership. This strategy contributes to the development of a lasting relationship between the organisation and its stakeholders.

Narrowing enterprise architecture. The Open Group Architecture Framework (TOGAF) is an all-encompassing EA ③ framework [45]. It offers a standardised terminology, methodology, and set of tools for designing, planning, and implementing enterprise architecture [45]. Four domains comprise TOGAF: business, data, application, and technology. These domains are interdependent and provide an enterprise-wide perspective. TOGAF is particularly relevant in the context of SGs, as they frequently involve the use of complex technology systems, and it is essential that these systems are designed and implemented to support the learning objectives. The TOGAF framework, moreover, can provide a structure for the development of SGs, ensuring that they cover all required topics and provide an understanding of concepts explored.

The EARF [46] definition of EA is accepted for the purposes of this research, as the ongoing process of defining the key parts of a sociotechnical organisation, as well as their interactions with others within the environment to come to grips with complexity and regulate change. TOGAF is a business architectural framework.

The pillars under investigation are narrowed and refined further in Table 3 below:

Table 3. Integrative literature review: Narrowing the pillars of study.

Pillar	Details
Serious games	**Serious Games (SGs) Definition and Components** • Essence of SGs: SGs combine entertainment with functional elements for user engagement across various sectors [47]. • Multimodal Media: They incorporate text, images, animations, audio, and haptics to deliver messages or skills [47]. • Purpose: The term 'serious' denotes the game's role in imparting knowledge, skill, or content [47]. • Experience: Player immersion in SGs relates to specific settings like education and health. • Components: Laamarti et al. [47] identify SGs as comprising experience, entertainment, and multimedia. **Academic and Industry Perspectives on SGs** SGs are deployed across various domains, including K-12 and higher education, healthcare, corporate sectors, the military, and non-governmental organizations, serving as versatile educational tools. They facilitate learning by providing media that enhance engagement and understanding, while also promoting the development of analytical skills and the application of knowledge in various scenarios. SGs are placed within the broader contexts of e-learning, edutainment, and game-based learning, indicating their wide applicability. They offer immersive environments for practicing skills and are recognised for their accessibility, enabling personalised learning that can be accessed at any time and place, thus supporting a more inclusive approach to education. **Critical SG Variables for Definition Development** • Game Purpose: Entertainment and learning [38]. • Format: Digital and non-digital [38]. • Genre: Includes action, adventure, puzzle, platforming, etc. [38]. • Platform/Delivery: Mobile, online, PC, game console, etc. [38]. • Application Area: Business, engineering, history, languages, etc. [38].
Stakeholder theory	**Stakeholder Theory (ST) Evolution and Classification** • Development: ST has evolved to differentiate and categorize stakeholders based on power, legitimacy, and urgency [48]. • ST Cataloguing Model: Miles [48] provides a model with four stakeholder types: ○ Influencer: Individuals with the ability to affect organizational actions. ○ Claimant: Individuals with a claim on the organization which lack the power to enforce it. ○ Collaborator: Individuals who cooperate with the organization without an active influence. ○ Recipient: Individuals impacted by organizational operations without active claim pursuit.
Enterprise architecture	**Enterprise Architecture (EA) Development and Functions** • Origins: EA emerged to address complex distributed technological systems [49]. • Significance: EA is crucial for navigating the complexities of the digital business landscape. • Transformation: EA establishes organizational patterns for facilitating change [50]. • Management: An EA team develops and manages the organization's framework, aligning with high-level business strategies. • Business Intelligence: EA role parallels a Business Intelligence architect, aiming to save time and costs through effective change management [50]. • Governance: EA requires governance frameworks to oversee the architecture and ensure alignment with organizational objectives [50]. **EA's Contribution to Organizational Functions** • Project Management: EA defines work packages integrated into project plans, helping to allocate resources and estimate timeframes [50]. • Service Design: EA patterns ensure service warranties are met [50]. • IT Governance: EA reviews and controls IT asset purchases to align with broader requirements and governance directions. • Process Improvement: EA acts as a repository for process improvement knowledge, enabling standards updates for efficiency gains [50]. **EA's Role in Business Process Management** • EA provides a structural framework supporting Business Process Management, Business Intelligence, Information Management, and Data Management.

Pillar	Details
	• Serious Games Variables: Game purpose, format, genre, platform/delivery, and application area are critical for defining SGs [38].

3.4 Framing the pillars of study

Framing stakeholder theory. Any organisation and consequent activities/projects will be better positioned for success if they can incorporate the views of and involve stakeholders to help shape them.

Framing enterprise architecture. Being equipped with the potential affordances of both SGs ① and ST ② is not enough. One would need governance structures, I.T. strategies, configuration items, project charters and more to successfully meet the goals of this research. A systematic literature review conducted by Gong and Janssen [51] reveals categories of value with related evidence from academic works for this research pillar. The discipline acts a tool—informed by EA ③ form and context—to handle complexity and generate value [51]

3.5 Methods in the pillars of study

Methods in stakeholder theory. Rabinowitz [52] describes stakeholder theory (ST) as a participatory process that involves all individuals impacted by a project to improve procedures and increase support. Identifying relevant stakeholders and anticipating challenges is crucial to harness the benefits of ST, which includes stakeholder identification, analysis, and management.

Stakeholder identification is an ongoing, dynamic process, essential to organizational practice, requiring iterative refinement as stakeholder dynamics evolve. The outcome of stakeholder identification is a comprehensive stakeholder list, regularly maintained for accuracy, detailing names, titles, contact information, and affiliations, which is foundational to stakeholder analysis. Schmeer [53] notes that stakeholder analysis systematically gathers and analyses information to determine whose interests should be considered in policy or program development. It involves assessing stakeholders' knowledge, interests, alliances, positions, and their capacity to influence policies through leadership or power. This analysis is crucial for management to engage stakeholders effectively, build support for policies, and pre-empt potential conflicts, thus facilitating successful policy or program implementation.

Stakeholder analysis involves the following steps [53]:

1. Planning: Define the analysis's purpose, identify the information's users, and develop plans and timelines for its utilization.
2. Policy Selection: Choose a specific issue or policy to focus on, providing clear and concise definitions.
3. Identifying Stakeholders: Analyse documents to gather relevant information, create a list of potential stakeholders, and refine this list to prioritize key stakeholders based on expert input.

4. Tool Adaptation: Specify the information needed from stakeholders, including their knowledge of the policy, position, interests, alliances, resources, power, and leadership capabilities. Develop and test an interview protocol.
5. Information Collection: Review existing information, schedule interviews, and gather data on stakeholders' positions, interests, and influence.
6. Stakeholder Table: Record stakeholders' positions and assess their power/interest levels.
7. Analysis: Perform analyses on power, leadership, knowledge, positions, interests, and alliances. Reflect on the gathered data.
8. Utilization: Use the analysed data to present results and inform stakeholder management strategies.

Methods in serious games. Warren and Jones [54] argue that no universal approach exists for SG development, likening it to an artform akin to curriculum design or lesson planning. A successful SG requires a careful balance of various elements, including instruction, educator support, student engagement, and other variables, both controllable and uncontrollable. The resulting framework should consider these complex variables to enhance the effectiveness of SG teams.

McGuire and Jenkins [55] observe that game development teams can range from solo developers to large corporations, with the team's size reflecting the project's scope and objectives. Core game development uses technology to realize game designs and mechanics conceptualized during the design phase, bridging technology and design. Development teams typically produce mechanics, content, and technology.

Weststar [56] suggests viewing game development stakeholders as an occupational community with unique identities, meanings, and aspirations. This perspective helps analyse their alignment or deviation from organizational, industry, and social norms.

SG development focuses on educational or training objectives, differing from traditional game development in several key aspects:

- Design Methodology: SGs prioritize educational goals over entertainment, often requiring collaboration with experts [57].
- Objectives: Gameplay mechanics in SGs are tailored to achieve specific learning outcomes, like promoting critical thinking [58].
- Evaluation: Assessing whether SGs meet their objectives is a critical part of development, involving effectiveness evaluations [59].
- Collaboration: SG development involves cooperation between developers, subject matter experts, educators, and other stakeholders [60].
- Target Audience: SGs may cater to specialized groups, influencing their design and development [61].

Overall, SG development demands a thoughtful and targeted approach to game design and development, with a strong emphasis on educational or training objectives.

Methods in enterprise architecture. The TOGAF Architecture Development Method (ADM) prescribes a sequence for architecture development but leaves scope determination to the organization, emphasizing the iterative nature of the process, where scope and deliverables evolve with each cycle [31]. Each ADM iteration contributes to the company's Architecture Continuum. A project's scope is pivotal for success, with complexity determined by the extent of horizontal and vertical integration as suggested by the Zachman Framework [45]. Regular validation against initial expectations is necessary throughout the ADM cycle [45]. The ADM's stages, such as the architecture development phases (B, C, and D), include steps like selecting reference models, tools, establishing baseline and target architecture descriptions, gap analysis, roadmap components definition, resolving Architecture Landscape effects, stakeholder reviews, completing the architecture, and composing the Architecture Definition Document [45].

The Requirements Management phase is continuous, ensuring changes in requirements are governed and reflected in all phases [45]. Organizations may use a Requirements Repository to track requirements, including those in the existing Statement of Architecture Work. Output from each stage is subject to modification in subsequent stages, with version control for output versioning.

4 Discussion

The stakeholder-centred framework for serious game (SG) design acts as a decision-support tool, addressing communication issues, focusing on relevant factors, and aiding in the evidence-based production of SG media. We therefore highlight multiple requirements for this framework: The framework for SG development should analyse SG teams' structures and conditions, identifying probable communication issues for evaluation throughout the project lifecycle. It should facilitate project participation and empowerment, leveraging human capital against technical requirements to ensure stakeholder effectiveness within their roles. Furthermore, the framework must accommodate project scope and objectives, advocating for a modular and extensible design to serve various stakeholders.

Changes in the system should be clearly communicated, adopting a common lexicon for a shared understanding among stakeholders, emphasising transparency. The focus should remain on creating effective SG media, not on the technology used, with the framework integrating smoothly with internal functions. The framework should, moreover, allow processes to be consistently replicated, optimising project components and simplifying variables like time and budget.

It must also be tailored to the people and relationships involved in SG design, valuing their connections and roles. Offering support to stakeholders is crucial, with a built-in support system to resolve issues and promote community engagement. The framework should be maintainable and upgradable, ensuring its longevity and reducing duplication of efforts. It should be ergonomic and user-friendly, simplifying the build and deployment processes. Ultimately, the framework must ensure high SG design standards, leading to productive media and institutionalising best practices.

SG development challenges are notable, including high production costs, difficulty in measuring effectiveness, multi-platform development demands, and maintaining player engagement [57,61,62].

ST and EA provide specific advantages in SG development:

ST Advantages:

- It structures the engagement of all stakeholders in the development process [63].
- It ensures a balanced consideration of all interests during decision-making [64].
- It enhances communication transparency and collaboration [65].
- It upholds ethical considerations, potentially increasing market acceptance [66].

EA Advantages:

- It offers a structured methodology for SG technical infrastructure, optimizing time and costs [67].
- It aligns SG development with organizational goals, ensuring strategic relevance [68].
- It improves user experience by focusing on end-user needs [69].
- It ensures flexibility and adaptability for future updates and market changes [70].

Combining ST and EA in SG development can lead to more engaging, effective, and strategically aligned games, considering the needs and interests of all parties involved, from players to investors and regulators. This approach can foster informed decision-making in design and marketing, enhancing player trust and aligning the game with broader social and environmental goals.

5 Limitations

Despite the comprehensive nature of our integrative literature review, this methodology does present some challenges. The review's validity hinges on the quality of the included studies as a convoluted research design or execution in the primary studies could undermine our findings. There is also potential for bias in the selection and interpretation of studies, as well as the risk of overlooking valuable non-English or unpublished research. Moreover, heterogeneity across studies in terms of design, settings, or populations complicates comparison and synthesis. Finally, given the fast-paced nature of SG research, some findings might become outdated by the time of publication. We have made every effort to mitigate these limitations in our approach.

6 Conclusion

This study has developed a comprehensive stakeholder-centred conceptual framework by integrating the domains of SGs, ST, and EA. By addressing the needs and interests of all stakeholders and aligning with organizational goals, this framework enhances SG development, ensuring both scalability and improved player experiences. By integrating ST [②] and EA [③] principles, SG designers [①] are equipped to develop a comprehensive strategy for games that meet stakeholder needs and expectations. Centring SG design on the multifaceted interests of stakeholders and considering their input becomes a priority. TOGAF's enterprise architecture provides a blueprint for aligning SG design with organizational business objectives, ensuring relevant and engaging learning experiences that improve performance and satisfaction.

A stakeholder-centred approach to SG development is crucial for creating effective, sustainable, and engaging games. Furthermore, EA offers a structured method for creating, implementing, and managing the game's technological infrastructure. This includes articulating technical specifications, selecting appropriate hardware and software platforms, and establishing development standards and best practices. Additionally, EA ensures that the game's technology is scalable, reliable, and secure, aligning with the overall business strategy and objectives.

This framework addresses several challenges in SG development. It facilitates communication and collaboration among developers, end-users, investors, and regulators [71], streamlines decision-making, aligns stakeholder goals, and ensures efficient resource allocation [72]. It emphasizes the player experience as a central stakeholder concern to maintain engagement and meet expectations [73]. Risk mitigation strategies address investor and regulator concerns about cost, compliance, and return on investment [74], while also considering stakeholders' technological compatibility needs [75].

In essence, the proposed stakeholder-centred framework for SG development emphasizes the importance of engaging all relevant stakeholders in a collaborative, unified effort. By doing so, it ensures that SGs are not only educational and engaging but also practical and relevant to the stakeholders' needs, as exemplified by the applications in student learning and physical therapy outlined earlier. The paper provides an in-depth exploration of the underlying literature that will inform the remainder of this project. With the framework's requirements and characteristics now established, future research will delve into system techniques and procedures, followed by a review of design choices by specialists in the field. The compatibility and value of the investigated domains forge a solid foundation for a stakeholder-centred SG design framework, which will undergo further maturation and scrutiny by experts for future implementation. The amalgamation of SGs, ST, and EA emerges as a robust strategy for developing effective and sustainable game-based solutions within higher education and other contexts.

Further research matures and visualises the conceptual framework from this integrative literature review and sees various experts examine and scrutinise the tool for future implementation. Ultimately, the combination of SGs, stakeholder theory, and enterprise architecture provides a promising strategy for developing effective and enduring game-based solutions for the context of higher education and beyond.

References

1. Kanode CM, Haddad HM. Software engineering challenges in game development. 2009 Sixth International Conference on Information Technology: New Generations. April 2009:260-265. doi:10.1109/ITNG.2009.74
2. Morgan G. Challenges of online game development: A review. Simul Gaming. 2009;40(5):688-710. doi:10.1177/1046878109340295
3. Rosyid HA, Palmerlee M, Chen K. Deploying learning materials to game content for serious education game development: A case study. Entertain Comput. 2018;26:1-9. doi:10.1016/j.entcom.2018.01.001

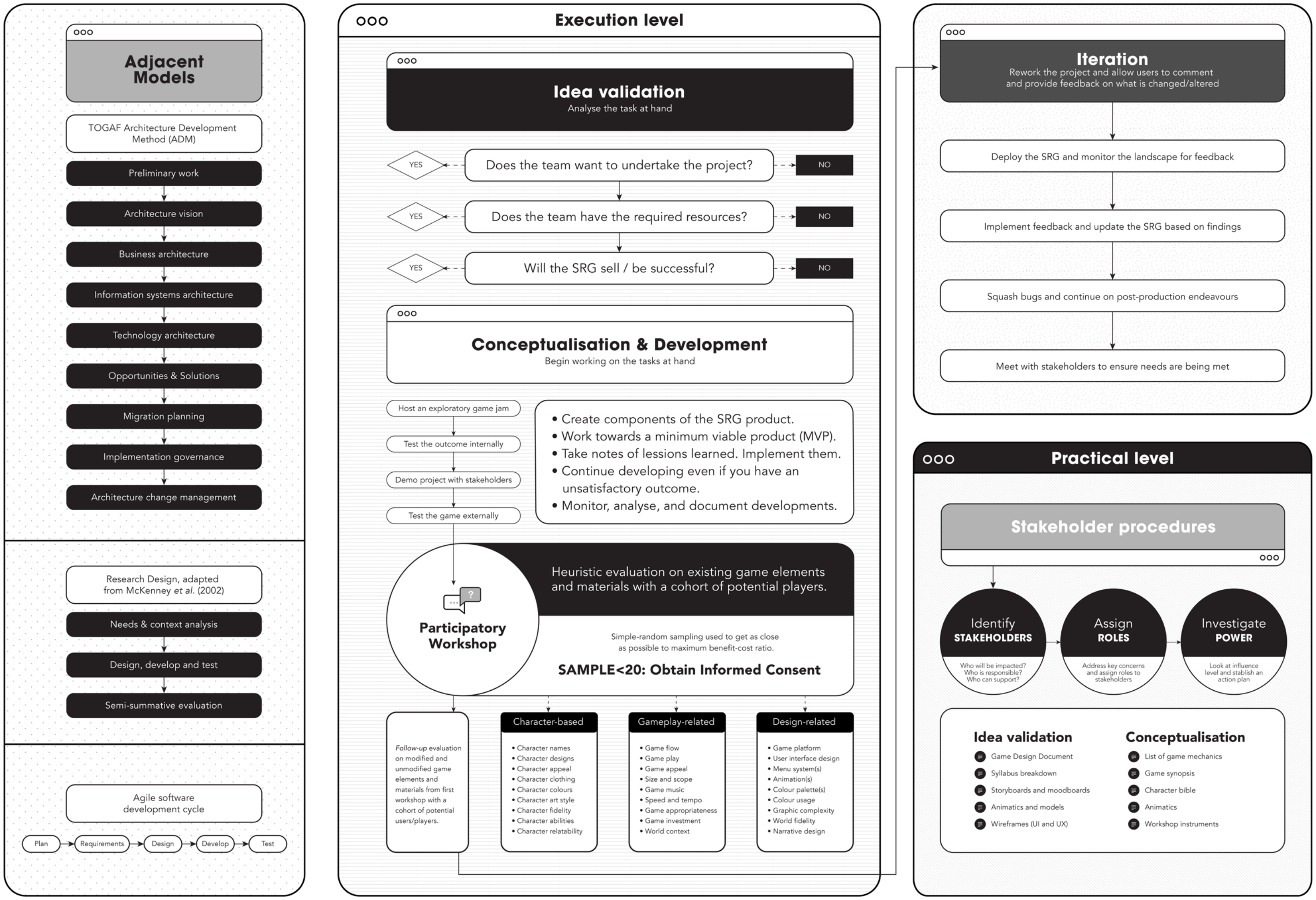
4. Whitson JR. What can we learn from studio studies ethnographies?: A “messy” account of game development materiality, learning, and expertise. Games Cult. 2020;15(3):266-288. doi:10.1177/1555412018783320
5. Flood S, Cradock-Henry NA, Blackett P, Edwards P. Adaptive and interactive climate futures: systematic review of ‘serious games’ for engagement and decision-making. Environ Res Lett. 2018;13(6):063005. doi:10.1088/1748-9326/aac1c6
6. Garneli V, Giannakos M, Chorianopoulos K. Serious games as a malleable learning medium: The effects of narrative, gameplay, and making on students’ performance and attitudes. Br J Educ Technol. 2017;48(3):842-859. doi:10.1111/bjet.12455
7. Guillén-Nieto V, Aleson-Carbonell M. Serious games and learning effectiveness: The case of It’sa Deal!. Comput Educ. 2012;58(1):435-448. doi:10.1016/j.compedu.2011.07.015
8. Hughs A. 11 Common Challenges You May Face During Serious Games Development. Game-based learning. elearningindustry.com. Published 2019. https://elearningindustry.com/serious-games-development-common-challenges [accessed Feb 16, 2023].
9. Michael DR, Chen SL. Serious Games: Games That Educate, Train, and Inform. Muska & Lipman/Premier-Trade; 2005. ISBN:1592006221
10. Bourne L. Making Projects Work: Effective Stakeholder and Communication Management. CRC Press; 2015. ISBN:1482206676
11. Marfisi-Schottman I, George S, Tarpin-Bernard F. Tools and methods for efficiently designing serious games. 4th Europeen Conference on Games Based Learning ECGBL. October 2010:226-234. ISBN:1906638780
12. Grzesiuk K. Network-driven approach to human resources management. Network. 2017;5(4):129-143. doi:10.29015/ceejme.648
13. Angafor GN, Yevseyeva I, He Y. Game-based learning: A review of tabletop exercises for cybersecurity incident response training. Secur Priv. 2020;3(6):e126. doi:10.1002/spy2.126
14. Edwards P, Sharma-Wallace L, Wreford A, et al. Tools for adaptive governance for complex social-ecological systems: a review of role-playing-games as serious games at the community-policy interface. Environ Res Lett. 2019;14(11):113002. doi:10.1088/1748-9326/ab4036
15. Caruso F, Peretti S, Santa Barletta V, Pino MC, Di Mascio T. Recommendations for developing Immersive Virtual Reality Serious Game for Autism: Insights from a Systematic Literature Review. IEEE Access. 2023. doi:10.1109/ACCESS.2023.3296882
16. Zwane Z, Matsiliza NS. Stakeholders’ involvement in service delivery at eDumbe Municipality. J Local Gov Res Innov. 2022;3:45. doi:10.4102/jolgri.v3i0.45
17. Lalicic L, Weber-Sabil J. Stakeholder engagement in sustainable tourism planning through serious gaming. In: Qualitative Methodologies in Tourism Studies. Routledge; 2022:192-212. ISBN:9781032227740
18. Jayasuriya S, Zhang G, Yang RJ. Exploring the impact of stakeholder management strategies on managing issues in PPP projects. Int J Constr Manag. 2020;20(6):666-678. doi:10.1080/15623599.2020.1753143
19. Zikargae MH, Woldearegay AG, Skjerdal T. Assessing the roles of stakeholders in community projects on environmental security and livelihood of impoverished rural society: A non-governmental organization implementation strategy in focus. Heliyon. 2022;8(10). doi:10.1016/j.heliyon.2022.e10987
20. Gulati R, Wohlgezogen F. Can purpose foster stakeholder trust in corporations?. Strategy Science. 2023. doi:10.1287/stsc.2023.0196
21. Kechagioglou P. Learning from Innovation Failure—A Case Study. In: Healthcare Innovation Success: Learning from Organisational Experience. Springer Nature Switzerland; 2023:41-82. doi:10.1007/978-3-031-28353-6_3

22. Angafor GN, Yevseyeva I, He Y. Game-based learning: A review of tabletop exercises for cybersecurity incident response training. Secur Priv. 2020;3(6):e126. doi:10.1002/spy2.126
23. Torraco RJ. Writing integrative literature reviews: Guidelines and examples. Hum Resour Dev Rev. 2005;4(3):356-367. doi:10.1177/1534484305278283
24. Alcayaga A, Wiener M, Hansen EG. Towards a framework of smart-circular systems: An integrative literature review. J Clean Prod. 2019;221:622-634. doi:10.1016/j.jclepro.2019.02.085
25. Onwuegbuzie AJ, Frels R. Seven Steps to a Comprehensive Literature Review: A Multimodal and Cultural Approach. SAGE Publications Ltd; 2016. ISBN:9781446248911
26. Szostak R. The interdisciplinary research process. In: Case Studies in Interdisciplinary Research. SAGE Publications Inc; 2012:3-19. ISBN:9781412982481
27. Freeman RE. Strategic Management: A Stakeholder Approach. Cambridge University Press; 2010. ISBN:9780521151740
28. Abt CC. Serious Games. University Press of America; 1987. ISBN:9780819161475
29. Freeman RE, Harrison JS, Zyglidopoulos S. Stakeholder Theory: Concepts and Strategies. Cambridge University Press; 2018. ISBN:9781108539500
30. Susi T, Johannesson M, Backlund P. Serious Games: An Overview. Technical Report HS-IKI -TR-07-001; School of Humanities and Informatics, University of Skövde, Sweden; 2007. https://www.diva-portal.org/smash/record.jsf?pid=diva2%3A2416&dswid=-5547 [accessed Feb 20, 2023].
31. Lankhorst M. Enterprise Architecture at Work. Springer: Berlin; 2017. doi:10.1007/978-3-642-29651-2
32. Gurbuz SC, Celik M. Serious games in future skills development: A systematic review of the design approaches. Comput Appl Eng Educ. 2022;30(5):1591-1612. doi:10.1002/cae.22557
33. Bunt LR. Investigating perceptions of stakeholders' positions, activities and specialisations at a serious game interest area (NWU, Vaal). Doctoral dissertation, North-West University (South Africa). 2020. http://hdl.handle.net/10394/34718 [accessed Feb 25, 2023].
34. Anthony Jnr B. Managing digital transformation of smart cities through enterprise architecture–a review and research agenda. Enterp Inf Syst. 2021;15(3):299-331. doi:10.1080/17517575.2020.1812006
35. Wehrle R, Wiens M, Schultmann F. Application of collaborative serious gaming for the elicitation of expert knowledge and towards creating Situation Awareness in the field of infrastructure resilience. Int J Disaster Risk Reduct. 2022;67:102665. doi:10.1016/j.ijdrr.2021.102665
36. Carrión-Toro M, Santorum M, Acosta-Vargas P, Aguilar J, Pérez M. iPlus a user-centered methodology for serious games design. Appl Sci. 2020;10(24):9007. doi:10.3390/app10249007
37. Puskaric H, Zahar Djordjevic M, Djordjevic A. Game development and connection to modern software engineering. 2023. https://scidar.kg.ac.rs/handle/123456789/18407 [accessed Feb 25, 2023].
38. Connolly TM, Boyle EA, MacArthur E, Hainey T, Boyle JM. A systematic literature review of empirical evidence on computer games and serious games. Comput Educ. 2012;59(2):6. doi:10.1016/j.compedu.2012.03.004
39. Gee JP. What video games have to teach us about learning and literacy. Comput Entertain. 2003;1(1):20-20. doi:10.1145/950566.950595
40. Djaouti D, Alvarez J, Jessel JP, Rampnoux O. Origins of serious games. In: Serious Games and Edutainment Applications. Springer; 2011:25-43. doi:10.1007/978-1-4471-2161-9_3

41. Graafland M, Schraagen JM, Schijven MP. Systematic review of serious games for medical education and surgical skills training. Br J Surg. 2012;99(10):1322-1330. doi:10.1002/bjs.8819
42. Papastergiou M. Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. Comput Educ. 2009;52(1):1-12. doi:10.1016/j.compedu.2008.06.004
43. Kapp KM. The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education. John Wiley & Sons; 2012. ISBN:1118096347
44. Lameras P. Essential features of serious games design in higher education. Learn. 2015;4(5). https://srhe.ac.uk/wp-content/uploads/2020/03/LamerasEssential_Features_of_-Serious_Games-Design_Short_FINAL.pdf [accessed Feb 20, 2023].
45. The Open Group Standard. The TOGAF® Standard, Version 9.2. https://pubs.opengroup.org/architecture/togaf9-doc/arch/ [accessed Feb 20, 2023].
46. EARF. EARF Enterprise Architecture Definition. Enterprise Architecture Research Forum. 2009. https://samvak.tripod.com/earf.pdf [accessed Feb 20, 2023].
47. Laamarti F, Eid M, Saddik AE. An overview of serious games. Int J Comput Games Technol. 2014:11. doi:10.1155/2014/358152
48. Miles S. Stakeholder theory classification, definitions and essential contestability. In: Stakeholder Management. Emerald Publishing Limited; 2017. doi:10.1108/S2514-175920170000002
49. Zachman JA. Concepts of the Framework for Enterprise Architecture. Los Angels, CA. 1996.
50. Borwick J. A Brief Introduction to Enterprise Architecture (EA). HEIT Management. 2013. http://www.heitmanagement.com/blog/2013/10/a-brief-introduction-to-enterprise-architecture/ [accessed Feb 20, 2023].
51. Gong Y, Janssen M. The Value of and Myths About Enterprise Architecture. Int J Inform Manage. 2019;46:1-9. doi:10.1016/j.ijinfomgt.2018.11.006
52. Rabinowitz P. Section 8: Identifying and Analyzing Stakeholders and Their Interests. Community Tool Box. ctb.ku.edu. Published 2021. https://ctb.ku.edu/en/table-of-contents/participation/encouraging-involvement/identify-stakeholders/main [accessed Feb 20, 2023].
53. Schmeer K. Stakeholder Analysis Guidelines. Policy Toolkit for Strengthening Health Sector Reform. 1999;1:1-35. https://cnxus.org/wp-content/uploads/2022/04/Stakeholders_analysis_guidelines.pdf [accessed Feb 24, 2023].
54. Warren SJ, Jones G. Learning Games: The Science and Art of Development. Springer; 2017. ISBN:3319468294
55. McGuire M, Jenkins OC. Creating Games: Mechanics, Content, and Technology. CRC Press; 2008. ISBN:1439865922
56. Weststar J. Understanding Video Game Developers as an Occupational Community. Inform Commun Soc. 2015;18(10):1238-1252. doi:10.1080/1369118X.2015.1036094
57. Ritterfeld U, Cody M, Vorderer P, eds. Serious Games: Mechanisms and Effects. Routledge; 2009. ISBN:1135848912
58. Ferdig RE, ed. Handbook of Research on Effective Electronic Gaming in Education. IGI Global; 2008. ISBN: 1599048116
59. Schrier K, Gibson D, eds. Designing Games for Ethics: Models, Techniques and Frameworks: Models, Techniques and Frameworks. IGI Global; 2010. ISBN:160960122X
60. Kankaanranta MH, Neittaanmäki P, eds. Design and Use of Serious Games. Vol 37. Springer Science & Business Media; 2008. ISBN:1402094965
61. Bellotti F, Kapralos B, Lee K, Moreno-Ger P, Berta R. Assessment in and of Serious Games: An Overview. Adv Hum-Comput Interact. 2013;2013:1-1. doi:10.1155/2013/136864

62. Wouters P, Van Nimwegen C, Van Oostendorp H, Van Der Spek ED. A Meta-Analysis of the Cognitive and Motivational Effects of Serious Games. J Educ Psychol. 2013;105(2):249. doi:10.1037/a0031311
63. Miller JA, Vepřek LH, Deterding S, Cooper S. Practical Recommendations from a Multi-Perspective Needs and Challenges Assessment of Citizen Science Games. PLoS ONE. 2023;18(5):e0285367. doi:10.1371/journal.pone.0285367
64. Street J, Stafinski T, Lopes E, Menon D. Defining the Role of the Public in Health Technology Assessment (HTA) and HTA-Informed Decision-Making Processes. Int J Technol Assess Health Care. 2020;36(2):87-95. doi:10.1017/S0266462320000094
65. Barrane FZ, Ndubisi NO, Kamble S, Karuranga GE, Poulin D. Building Trust in Multi-Stakeholder Collaborations for New Product Development in the Digital Transformation Era. Benchmarking. 2021;28(1):205-228. doi:10.1108/BIJ-04-2020-0164
66. Waheed A, Zhang Q. Effect of CSR and Ethical Practices on Sustainable Competitive Performance: A Case of Emerging Markets From Stakeholder Theory Perspective. J Bus Ethics. 2022;175(4):837-855. doi:10.1007/s10551-020-04679-y
67. Jnr BA, Petersen SA. Validation of a Developed Enterprise Architecture Framework for Digitalisation of Smart Cities: A Mixed-Mode Approach. J Knowl Econ. 2023;14:1702–1733. doi:10.1007/s13132-022-00969-0
68. Aldea A, Vaicekauskaitė E, Daneva M, Piest JPS. Assessing Resilience in Enterprise Architecture: A Systematic Review. In: 2020 IEEE 24th International Enterprise Distributed Object Computing Conference (EDOC). IEEE; 2020:1-10. doi:10.1109/EDOC49727.2020.00011
69. Van Wessel RM, Kroon P, De Vries HJ. Scaling Agile Company-Wide: The Organizational Challenge of Combining Agile-Scaling Frameworks and Enterprise Architecture in Service Companies. IEEE Trans Eng Manage. 2021;69(6):3489-3502. doi:10.1109/TEM.2021.3128278
70. Шабан АП, Биккулова ЗУ. How to Use Enterprise Modeling and Enterprise Architecture to Evaluate and Demonstrate the Value and Impact of Digital Innovation (or Other Types of Implementation of IT)?. Published 2020. https://elibrary.ru/item.asp?id=49930831 [accessed Feb 24, 2023].
71. Duc AN, Abrahamsson P. Minimum Viable Product or Multiple Facet Product? The Role of MVP in Software Startups. In: Agile Processes, in Software Engineering, and Extreme Programming: 17th International Conference, XP 2016, Edinburgh, UK, May 24-27, 2016, Proceedings 17. Springer International Publishing; 2016:118-130. doi:10.1007/978-3-319-33515-5_10
72. Phillips LD, Bana e Costa CA. Transparent Prioritisation, Budgeting and Resource Allocation with Multi-Criteria Decision Analysis and Decision Conferencing. Ann Oper Res. 2007;154(1):51-68. doi:10.1007/s10479-007-0183-3
73. Isbister K, Schaffer N. Game Usability: Advancing the Player Experience. CRC Press; 2008. ISBN:1498759572
74. Hardy G. Using IT Governance and COBIT to Deliver Value With IT and Respond to Legal, Regulatory and Compliance Challenges. Inform Secur Tech Rep. 2006;11(1):55-61. doi:10.1016/j.istr.2005.12.004
75. Pandey D, Suman U, Ramani AK. An Effective Requirement Engineering Process Model for Software Development and Requirements Management. In: 2010 International Conference on Advances in Recent Technologies in Communication and Computing. IEEE; 2010:287-291. doi:10.1109/ARTCom.2010.24

Annexure A: Conceptual Framework Rendering



Adjacent Models

TOGAF Architecture Development Method (ADM)

- Preliminary work
- Architecture vision
- Business architecture
- Information systems architecture
- Technology architecture
- Opportunities & Solutions
- Migration planning
- Implementation governance
- Architecture change management

Research Design, adapted from McKenney *et al.* (2002)

- Needs & context analysis
- Design, develop and test
- Semi-summative evaluation

Agile software development cycle

Plan → Requirements → Design → Develop → Test

Execution level

Idea validation

Analyse the task at hand

- Does the team want to undertake the project? YES / NO
- Does the team have the required resources? YES / NO
- Will the SRG sell / be successful? YES / NO

Conceptualisation & Development

Begin working on the tasks at hand

- Host an exploratory game jam
- Test the outcome internally
- Demo project with stakeholders
- Test the game externally

- Create components of the SRG product.
- Work towards a minimum viable product (MVP).
- Take notes of lessions learned. Implement them.
- Continue developing even if you have an unsatisfactory outcome.
- Monitor, analyse, and document developments.

Participatory Workshop

Heuristic evaluation on existing game elements and materials with a cohort of potential players.

Simple-random sampling used to get as close as possible to maximum benefit-cost ratio.

SAMPLE<20: Obtain Informed Consent

Follow-up evaluation on modified and unmodified game elements and materials from first workshop with a cohort of potential users/players.

Character-based	Gameplay-related	Design-related
Character names	Game flow	Game platform
Character designs	Game play	User interface design
Character appeal	Game appeal	Menu system(s)
Character clothing	Size and scope	Animation(s)
Character colours	Game music	Colour palette(s)
Character art style	Speed and tempo	Colour usage
Character fidelity	Game appropriateness	Graphic complexity
Character abilities	Game investment	World fidelity
Character relatability	World context	Narrative design

Iteration

Rework the project and allow users to comment and provide feedback on what is changed/altered

- Deploy the SRG and monitor the landscape for feedback
- Implement feedback and update the SRG based on findings
- Squash bugs and continue on post-production endeavours
- Meet with stakeholders to ensure needs are being met

Practical level

Stakeholder procedures

- Identify **STAKEHOLDERS**: Who will be impacted? Who is responsible? Who can support?
- Assign **ROLES**: Address key concerns and assign roles to stakeholders
- Investigate **POWER**: Look at influence level and stablish an action plan

Idea validation	Conceptualisation
Game Design Document	List of game mechanics
Syllabus breakdown	Game synopsis
Storyboards and moodboards	Character bible
Animatics and models	Animatics
Wireframes (UI and UX)	Workshop instruments

Strategies for Healthcare Resilience: A Comparative Evaluation of ARIMA and LSTM Models in Predicting COVID-19 Hospital Admissions

Taurai T. Chikotie [1,2*][0009-0008-6683-4987], Bruce Watson[1][0000-0003-0511-1837], Liam R. Watson[3][0000-0002-7016-9229], and Takudzwa Vincent Banda[1][0009-0003-2900-5518]

[1] Centre for AI Research (CAIR), Stellenbosch University, Cape Town, South Africa
[2] Department of Information Science, Stellenbosch University, Cape Town, South Africa
[3] Computer Science Department, University of Waterloo, Ontario, Canada

*Corresponding author(s). E-mail(s): taurai.chikotie@icloud.com.
Contributing authors: dr.bruce.watson@gmail.com; liamrwatson03@gmail.com; tadiwanashe-banda74@gmail.com

Abstract. The Eastern Cape Province faced significant challenges in hospital bed planning due to high COVID-19 infection rates and a lack of effective estimation models to support decision-making. This was because the first wave was characterised by the ancestral strain with a mutation at position Asp614Gly, the second by the beta variant (B.1.351), the third by the delta variant (B.1.617.2), and the fourth by the omicron variant (B.1.1.529). Therefore, there is a need to develop adequate short-term prediction models for forecasting the number of hospital admissions. This study proposed a comparative analysis of univariate (Autoregressive Integrated Moving Average) ARIMA, (Long-term Short Memory) LSTM, and ensemble ARIMA-LSTM models for COVID-19 hospital admissions forecasting. Leveraging historical data, we evaluated model performance in the public and private sectors using (Root Squared Mean Error) RSME, (Mean Absolute Error) MAE, and R-squared error. Our findings revealed that the LSTM model is the better performing model for both the public - RMSE: 0.146963, MAE: 0.1018066, R^2: 0.9990176 and private sectors - RMSE: 4.2412125, MAE: 0.0816642, R^2: 0.9990707. LSTM has the lowest RMSE and MAE values, indicating more accurate predictions, and the highest R^2 values, indicating the best fit to the actual data. The ensemble ARIMA-LSTM improves performance over the ARIMA model but is still not as good as the LSTM model. The ARIMA model has the poorest performance among the three models for both sectors. These models were also compared with other existing models from previous related studies. Due to its proven resilience and heightened predictive precision, using LSTM holds promise for enhancing pandemic forecasting, thereby facilitating improved hospital admissions planning and management strategies.

Keywords: Resilient Health Systems, Temporal Dependencies, South Africa

1 Introduction

Ever since its declaration as a pandemic by the World Health Organisation (WHO) in March 2020, the novel coronavirus disease (COVID-19) posed unparalleled challenges to global healthcare systems, marked by rapid escalation in infection and mortality rates within a condensed timeframe [1]. South Africa, a developing nation characterised by a bifurcated healthcare framework encompassing both public and private sectors, did not escape the relentless wave of challenges that enveloped the global community, as it emerged as one of the most profoundly impacted countries on the African continent [2-4].

Caught in a quagmire of all the health systems shock due to high infection and admission rates, hospital preparedness was found lagging as there were no specific planning guidelines in place to arrest the rampant increases in hospital admissions. A high number of hospital admissions meant high chances of strained health resources, compounded risks of more transmissions, and work overload to healthcare providers. Because of the inadequacy in the ability to use precision estimation methodologies, there were huge expenses were incurred in building mobile health facilities, buying protective equipment, and other health technologies [4-7].

Models such as ARIMA, and LSTM model have been previously employed in healthcare-related studies such as disease surveillance to assist in time series forecasting providing for better decision-making [8-11]. Using ARIMA and LSTM models for forecasting in healthcare admissions offers several advantages. ARIMA, a classical time series model, is particularly useful for capturing linear trends and seasonality in hospital admissions data [12]. Its simplicity and interpretability make it a valuable tool for short-term forecasting, allowing hospitals to allocate resources effectively. On the other hand, LSTM, a type of recurrent neural network (RNN), excels at capturing complex, nonlinear patterns in time series data, which can be especially beneficial in healthcare [13]. LSTM's ability to remember past information over long sequences and adapt to changing patterns makes it suitable for longer-term and more intricate hospital admissions forecasts.

Therefore, there is a need for developing adequate short-term prediction models for forecasting the number of hospital admissions. In this study, we proposed a comparative analysis of univariate ARIMA, LSTM, and ensemble ARIMA-LSTM models for COVID-19 hospital admissions forecasting. The rest of the paper unfolds as follows: In Section 2, conduct a comprehensive literature review, delving into existing research on the use of ARIMA, LSTM and other models in the prediction of COVID-19. Section 3 outlines the data collection and preprocessing methods, introduces the proposed models, and how the prediction process was conducted. Section 4 presents experimental results, model comparisons, and implications. Section 5 presents study discussions. Section 6 presents our concluding observations, summarising the key takeaways from the study.

2 Literature Review

The Eastern Cape Province was also engulfed in hospital bed planning challenges due to high infection rates and insufficient models of estimation to aid decision-making. Fig.1 summarises the trends in admission in the Eastern Cape. A timeline on the spread of the virus suggests that, like the National trend, between the period of March 2020 to September 2022, the Eastern Cape Province experienced four distinct waves of the COVID-19 pandemic, each primarily characterised by the prevalence of a specific variant of concern [14]. These waves featured the ancestral strain with a mutation at position Asp614Gly during the first wave, the beta variant (B.1.351) in the second wave, the delta variant (B.1.617.2) during the third wave, and finally, the omicron variant (B.1.1.529) in the fourth wave.

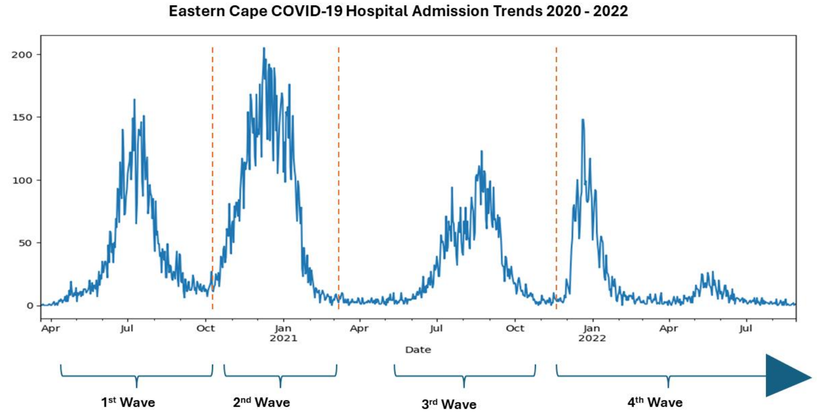


Fig. 1. Eastern Cape Province COVID-19 hospital admissions trends for the period 2020-2022

Table 1 presents models reported in the literature for forecasting Covid-19 pandemic-related studies. Most of the studies reported CNN, LSTM and Bi-LSTM models as outperforming statistical traditional models such as ARIMA and SVR.

Table 1. Models reported in the literature for forecasting Covid-19 pandemic-related studies.

Authors	Proposed Models	Remarks
Ma et al. [15]	LSTM, LSTM-Markov hybrid model	This study reported the comparison between LSTM and LSTM-Markov model but did not discuss the parameters of these models.
Nabi et al. [16]	CNN, Multivariate CNN	They reported that CNN has performed better than LSTM.
Ketu et al. [17]	CNN-LSTM, LSTM, ARIMA	The proposed hybrid CNN-LSTM model performed better than ARIMA or LSTM models. This study also lacks the information on the parameters of the models.
Shahid et al. [18]	ARIMA, Bi-LSTM, GRU, SVR	The findings show that the Bi-LSTM model consistently outperforms the others, achieving the lowest error rates and highest accuracy, making it a valuable tool for pandemic prediction and public health planning.
Valente et al. [19]	LSTM, Bi-LSTM,	The findings demonstrate state-of-the-art accuracy in ICU predictions, suggesting that these models can significantly aid policymakers in optimising public spending and managing regional mobility during the pandemic.
Devaraj et al. [20]	Stacked LSTM, ARIMA, LSTM, Prophet Approaches	This paper explores the use of deep learning-based time series techniques to predict COVID-19 cases, emphasising the importance of AI in forecasting pandemic trends. Among the models evaluated, Stacked LSTM showed the highest accuracy with less than 2% error, proving its reliability for predicting cumulative confirmed, death, and recovered cases globally, as well as in specific regions like India and Chennai.
Seo et al. [21]	Bi-LSTM,	This study aimed to create a web-based tool for hospital administrators to accurately predict bed occupancy rates (BOR) for individual wards and rooms over different time periods, enhancing hospital scheduling and resource management. The findings showed that Bi-LSTM models provided the best performance, with the ward-level model achieving a mean absolute error (MAE) of 0.067 and the room-level model combining dynamic and static data achieving an MAE of 0.129, facilitating efficient bed operation planning though visual web-based dashboards.

3 Method

3.1 Data Collection and Preprocessing

This study employed retrospective data comprising 46,050 records, with approximately 30,920 records from the public sector and around 15,130 records from the private sector, sourced from the South African DATCOV national active surveillance system for COVID-19 hospital admissions spanning from March 2020 to August 2022. Notably, the higher patient volume in the public sector can be attributed to its provision of predominantly free healthcare services, whereas the private sector reported fewer hospital admissions due to its reliance on medical insurance for service coverage. The dataset encompassed a total of 893 days' worth of information.

In the data preprocessing phase, we began by identifying the unique categories present in the "Sector" column of our dataset. This step is essential for understanding the categorical variables we are working with. We then utilised a common technique called one-hot encoding to transform these categorical variables into a numerical format that machine learning algorithms can interpret more effectively. One-hot encoding essentially creates binary columns for each category, indicating its presence or absence in each observation. This method prevents the algorithm from interpreting categorical variables as ordinal or numerical, which could lead to incorrect assumptions [22].

Following this, we focused on time series data manipulation. We converted the "Date" column into a datetime format to facilitate time-based operations. Converting dates to datetime format enables time-based analysis, allowing us to uncover temporal patterns and trends in the data. Aggregating data over specific time periods and computing moving averages provides insights into long-term trends while smoothing out short-term fluctuations, resulting in more robust analyses.

Finally, we computed the 14-day moving average of each sector’s hospitalisation admissions (see Fig. 2). Using a 14-day moving average helps to smooth out short-term fluctuations in the data, providing a clearer view of the underlying trend. This technique also helps to highlight longer-term patterns by averaging out daily variations, making it easier to discern gradual changes over time.

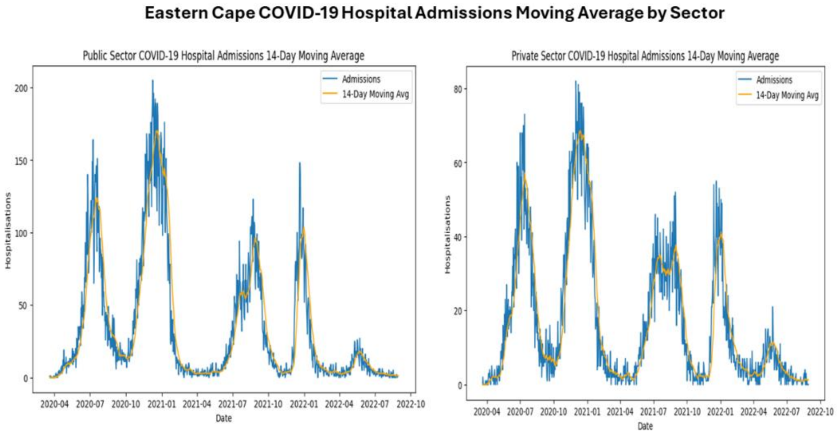


Fig. 2. EC COVID-19 14-Day Moving Average by Sector

For LSTM, we allocated 80% of the data for training and 20% for testing the models and prediction. This means that we utilised a dataset spanning 800 days, while for model testing, we employed a dataset spanning 93 days. For ARIMA we decomposed the time series data as this aided us in understanding the data's structure, achieving stationarity, accounting for seasonality, improving forecast accuracy, and performing effective diagnostic checks. The AD Fuller test was conducted for both sectors to test for stationarity. Both data sets had p-values less than 5% suggesting that the dataset was stationary. Prediction for both models sought to estimate the hospital admissions figure in the next 28 days of the dataset for better hospital bed planning.

3.2 Models

To facilitate a rigorous forecasting, we utilised the autoregressive integrated moving average (ARIMA) and long short-term memory (LSTM) models, employing these methodologies to discern the superior model in predicting hospital bed occupation through admissions within the respective sectors under consideration.

Autoregressive Integrated Moving Average (ARIMA)

This is a popular time series forecasting model used to analyse and forecast data points based on their historical patterns including in healthcare-related studies [23]. ARIMA models are particularly effective for univariate time series data, where each data point is associated with a specific time index. Three main components constitute the ARIMA (p,d,q) model namely: Autoregressive (AR) terms, Integrated (I) order, and Moving Average (MA) terms. Each component captures different aspects of the time series data. The ARIMA model equation can be expressed as follows:

$y(t) = c + \varphi_1 y(t\text{-}1) + \varphi_2 y(t\text{-}2) + ... + \varphi_p y(t\text{-}p) + \theta_1 \varepsilon(t\text{-}1) + \theta_2 \varepsilon(t\text{-}2) + ... + \theta q \varepsilon(t\text{-}q) + \varepsilon(t)$

Where: y(t) represents the value of the time series at time t.

c is a constant term.

$_1$, φ_2, ..., φ_p are the autoregressive coefficients for the past p values of the series.

θ_1, θ_2, ..., θq are the moving average coefficients for the past q error terms (residuals).

$\varepsilon(t)$ is a white noise error term at time t.

Considering that this study employed the ARIMA model for its COVID-19 hospital admission forecasting analysis, a brief contextual explanation of the formula is as follows:

y(t): This represents the number of COVID-19 hospital admissions at time t.

φ_1, φ_2, ..., φ_p: These coefficients represent the autoregressive terms. In the context of COVID-19 hospital admissions, they capture the influence of past hospital admission numbers on the current hospital admissions count. For instance, if φ_1 is positive, it means that an increase in hospital admissions in the previous time step will positively impact the current hospital admissions count [23].

θ_1, θ_2, ..., θq: These coefficients represent the moving average terms. They reflect how past prediction errors (residuals) contribute to the current hospital admissions count. If θ_1 is positive, it indicates that past underpredictions (negative residuals) will contribute to higher hospital admission counts in the present.

$\varepsilon(t)$: This term represents the white noise error at time t. It captures the unexplained variability in hospital admissions that is not accounted for by the autoregressive and moving average terms. In the context of COVID-19 hospital admissions, $\varepsilon(t)$ could represent factors such as sudden outbreaks, policy changes, or other unforeseen events affecting hospital admission numbers.

c: The constant term captures any overall average or baseline level of hospital admissions that is not explained by the autoregressive and moving average components.

By fitting an ARIMA model to the historical COVID-19 hospital admissions data, we were able to estimate the values of φ_1, θ_1, and other coefficients to create a model that predicts future hospital admission numbers based on their historical patterns.

The ARIMA model for public sector hospital admissions possessed the following parameters: p = 5, d = 1, and q = 3. The ARIMA model for private sector hospital admissions possessed the following parameters: p = 3, d = 1, and q = 4. A comprehensive explanation of the ARIMA results is as follows:

Public Sector Hospital Admissions: ARIMA (5, 1, 3):

- p = 5: This indicates the number of autoregressive terms. It means the current value of the series is based on its previous two values. This suggests a relationship where the last two observations have a linear influence on the next observation.
- d = 1: This is the degree of differencing required to make the series stationary. A value of 1 means that the series needed to be differenced once to achieve stationarity. Differencing helps in removing trends or seasonal patterns, making the data's mean and variance constant over time.
- q = 3: This denotes the number of lagged forecast errors in the prediction equation, known as the moving average component. A q of 1 means the model uses the error term from the immediate last forecast to improve accuracy.

Private Sector Hospital Admissions: ARIMA (3, 1, 4):

- p = 3: Indicates that the current value of the series is influenced by its immediately previous value, showing a direct linear relationship with the last observation.
- d = 1: Like the public sector model, this series also required differencing once to achieve stationarity, indicating a trend that needed to be removed for effective modelling.
- q = 4: This suggests a more complex moving average component than in the public sector model. The current value of the series is influenced by the forecast errors from the previous three observations, implying that the prediction can be significantly adjusted based on the recent errors in forecasting.

In summary, the public sector model suggests that it relies on more of the immediate past values and less on the moving average component. This indicates a somewhat smoother series where recent past values are significant predictors of future values, with less noise. The private sector model, with a higher q value, suggests that the prediction error has a more pronounced effect on forecasting. This means that the private sector hospital admissions data is more volatile or has more noise, requiring a stronger moving average component to achieve accurate forecasts.

Both models requiring a value d of 1 suggest that each series had a trend component that was removed by differencing once, making them stationary and suitable for ARIMA modelling. The differences in p and q values between the two sectors highlight the unique characteristics of hospital admissions in these sectors, possibly due to varying factors influencing admissions in public versus private hospitals.

Long Short-term Memory

LSTM is a type of recurrent neural network (RNN) architecture that excels in capturing long-range dependencies and patterns in sequential data, making it particularly suitable for time series forecasting tasks [24, 25]. Unlike traditional feedforward neural networks, LSTM networks possess memory cells and specialised gating mechanisms that allow them to remember and update information over extended sequences. This unique architecture helps LSTMs avoid the vanishing gradient problem, which can hinder the training of deep networks on sequential data. As illustrated in Fig. 2, the primary components of an LSTM unit include:

Cell State (Ct): This is the core memory of the LSTM. It runs straight down the entire chain of LSTM units, allowing information to flow along the entire sequence with minimal modification. The cell state can be seen as a conveyor belt that runs through the unit with various gates controlling the flow of information.

Input Gate (i): This gate decides which information should be stored in the cell state. It uses the current input and the previous hidden state (output) to determine what new information to store.

Forget Gate (f): The forget gate decides what information should be removed from the cell state. It takes the previous hidden state and the current input to determine what information is no longer relevant.

Output Gate (o): The output gate decides what the next hidden state (output) should be. It combines the current input and the previous hidden state to produce the output.

Fig. 3.

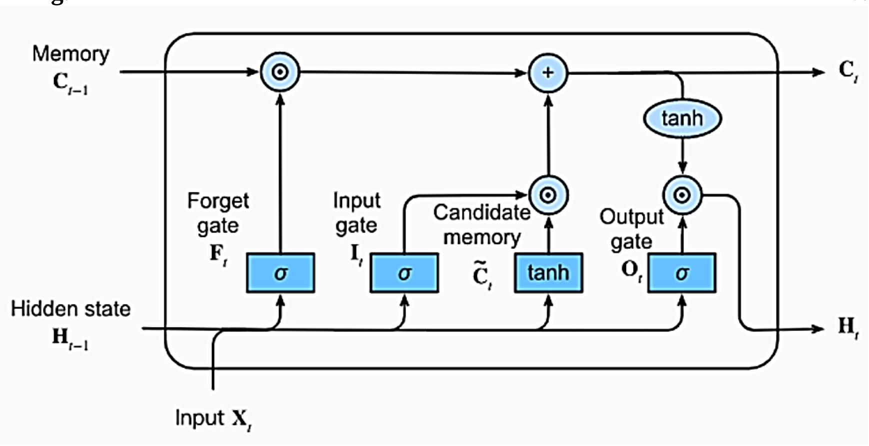


Fig. 4. Long Short-term Memory Architecture [17]

By applying LSTM to COVID-19 hospital admissions data, we were able to leverage its ability to capture intricate temporal relationships, adapt to changing dynamics, and make accurate forecasts.

We define a Long Short-Term Memory (LSTM) neural network model using the Keras Sequential API. The model architecture consists of three LSTM layers, each comprising 64 units, which are neural network cells capable of retaining information over time. The Rectified Linear Unit (ReLU) activation function is applied to each LSTM layer, introducing non-linearity to the model. Dropout layers, with a dropout rate of 20%, are added after each LSTM layer to prevent overfitting by randomly setting a fraction of input units to zero during training. This regularisation technique improves the model's ability to generalise to unseen data. A single neuron Dense layer follows the LSTM layers, serving as the output layer responsible for producing predictions. The model is compiled using the Adam optimiser and is trained to minimise the mean squared error (MSE) loss function. This loss function quantifies the disparity between the predicted and actual values, guiding the optimisation process during training. Overall, this model architecture is tailored for sequence prediction tasks, leveraging LSTM's memory capabilities, and incorporating dropout regularisation to enhance performance and prevent overfitting.

Ensemble ARIMA-LSTM

The ensemble model combines predictions from ARIMA and LSTM, by averaging their forecasts. This approach leverages the complementary strengths of both models to potentially improve overall prediction accuracy. By aggregating predictions from diverse modelling techniques, the ensemble model aims to mitigate individual model biases and capture a more robust representation of the underlying patterns in the data. In the study, this ensemble approach is motivated by the desire to enhance the accuracy of COVID-19 hospital admission forecasts by integrating the predictive capabilities of both traditional time series (ARIMA) and deep learning (LSTM) models. In addition, stacking method was also used but did not yield results. Table 2 summarises hyperparameters and values used in optimising the performance of the models in the study.

Table 2. Model's Hyperparameters and Value

Models	**Hyperparameter**	**Values**
LSTM	Number of neurons	50,100,50
	Epoch	20
	Verbose	0
	Batch size.	1
	Optimiser	Adam
	Loss function	MSE
	Activation function	ReLu
ARIMA	p, d, q	Public (5,1, 3), Private (3,14)
Ensemble ARIMA-LSTM	Combined LSTM and ARIMA	All LSTM and ARIMA values
	Weighted average	-

Model Evaluation

In the context of our study on forecasting hospital admissions in both public and private sectors, we have carefully selected RMSE, MAE, and R^2 as our primary evaluation metrics. RMSE represents the square root of the average of squared differences between predicted and observed values, emphasising larger errors. In contrast, MAE represents the average of absolute differences between predicted and observed values, providing insights into average prediction accuracy regardless of error direction. R^2 represents the proportion of the variance in the dependent variable that is predictable from the independent variables, and it is important because it indicates the goodness of fit and explanatory power of the mode [26]. RMSE, MAE and R^2 were chosen as evaluation metrics due to their simplicity, robustness to outliers, and alignment with common loss functions used in regression tasks [27]. These metrics offer intuitive interpretations and facilitate direct optimisation of models during training [28]. We aim to provide a concise yet comprehensive evaluation framework that enables meaningful comparisons between ARIMA, LSTM and ensemble ARIMA-LSTM models in predicting hospital admissions across public and private sectors. This approach ensures a fair and rigorous assessment of predictive performance, enhancing the reliability of our analyses in healthcare forecasting.

4 Experiment Results and Interpretation

4.1 ARIMA

Public Sector

For public sector hospital admissions, the ARIMA model yielded an RMSE of 8.6117031, an MAE of 6.9453619, and an R^2 of -2.3732639. These values suggest that, on average, the predictions are off by approximately 8.6117 units. The negative R^2 indicates that the model performs worse than a simple mean-based prediction, suggesting it does not capture the underlying trend well.

Private Sector

For private sector hospital admissions, the ARIMA model produced an RMSE of 6.7424687, an MAE of 5.7287073, and an R^2 of -3.8802909. These results indicate that the predictions are, on average, off by about 6.7425 units. The highly negative R^2 value again indicates poor model performance, signifying that the model fails to explain the variance in the data effectively.

4.2 LSTM

Public Sector

For public sector hospital admissions, the LSTM model achieved an RMSE of 0.1469631, an MAE of 0.1018066, and an R^2 of 0.9990176. These metrics demonstrate that the LSTM model provides highly accurate predictions, with errors averaging only 0.147 units. The near-perfect R^2 value indicates an excellent fit to the actual data, explaining almost all the variance.

Private Sector

For private sector hospital admissions, the LSTM model resulted in an RMSE of 0.0930407, an MAE of 0.0816642, and an R² of 0.9990707. This performance suggests very high prediction accuracy, with an average error of just 0.093 units. The R² value close to 1 confirms that the model almost perfectly captures the data's variance.

4.3 Ensemble ARIMA-LSTM

Public Sector

For public sector hospital admissions, the ensemble model combining ARIMA and LSTM produced an RMSE of 4.2412125, an MAE of 3.4279851, and an R² of 0.1818135. These results indicate that the ensemble model improves upon the ARIMA model but still falls short of the LSTM model in terms of accuracy and explanatory power. The positive R² suggests some degree of variance explanation, but it remains relatively modest.

Private Sector

For private sector hospital admissions, the ensemble model achieved an RMSE of 3.4005666, an MAE of 2.88388, and an R² of -0.2413962. These results show better performance than the ARIMA model alone but not as effective as the LSTM model. The negative R² value indicates that, despite some improvements, the ensemble model does not adequately capture the data's variance compared to the LSTM model.

Table 3 shows the results from the ARIMA, LSTM, and ensemble ARIMA-LSTM models between public and private sectors, and Fig 3 and Fig 4 depict public and private sector hospital admissions of all models against actual values.

Table 3. Results of the models' performance

Sector	Models	RSME*	MAE*	$R^{2\,*}$
Public	ARIMA	8.612	6.945	-2.373
	LSTM	0.147	0.102	0.999
	Ensemble ARIMA-LSTM	4.241	3.428	0.182
Private	ARIMA	6.742	5.729	-3.880
	LSTM	0.093	0.082	0.999
	Ensemble ARIMA-LSTM	3.401	2.884	0.241

* Rounded off to three decimal places.

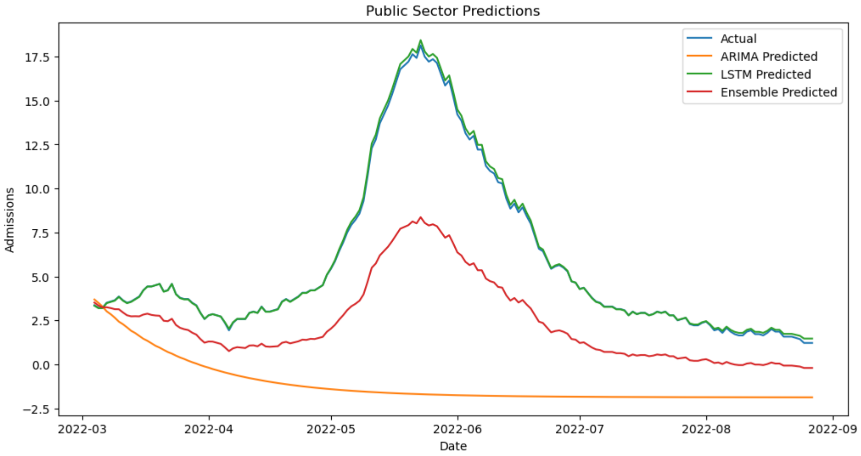


Fig. 4. Public sector predictions

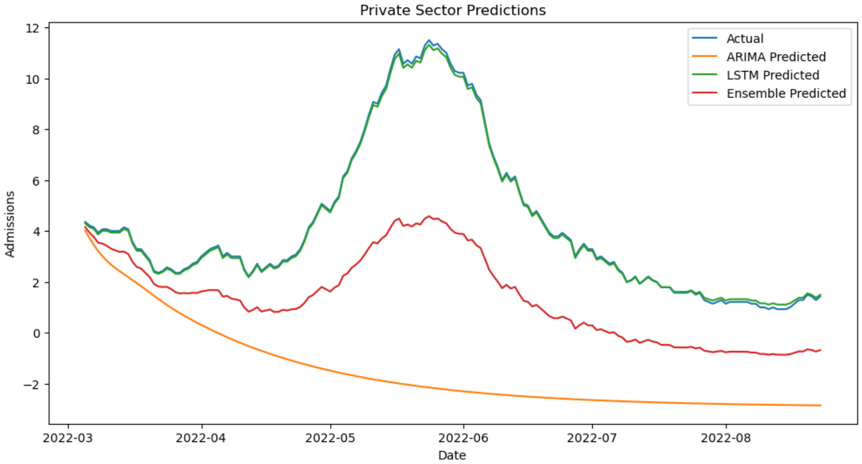


Fig. 5. Private sector predictions

4.4 Models Comparison Against Other Existing Models

The results of the ARIMA, LSTM, and ensemble ARIMA-LSTM were compared to those of previous studies despite them utilising different COVID-19 dataset and different geographical location. This comparison provided insights into the models' generalisability and performance across varying data sets and geographic locations. For comparison purpose we used the average score of public and private sectors for each model and metrics.

ARIMA Model:
ARIMA model shows an RMSE of 7.6770859, which is higher compared to most models in the previous research, indicating less accurate predictions. The MAE of 6.3370346 is also relatively high. The R² value is negative of -3.1267774, indicating that the model fits the data poorly. This aligns with previous research [17, 18, 20] findings where ARIMA models generally performed worse compared to deep learning models, such as the significantly negative R² values for various countries.

LSTM Model:
LSTM model demonstrates exceptional performance with an RMSE of 0.1200019 and MAE of 0.0917354, significantly lower than the values from the previous studies in [18]. For instance, in the UK, the LSTM model had a RMSE of 0.065. The R² value of 0.99904415 in the study suggests an almost perfect fit, which is much better than the results for Brazil (RMSE: 0.197) and Russia (RMSE:0.05) in [18]. This suggests that our LSTM model outperforms those in the previous research, especially in terms of accuracy and predictive power.

Ensemble Model:
The ensemble model in the study shows an RMSE of 3.82088955, which is higher than the LSTM in [15,17] but lower than the ARIMA model [18], suggesting moderate performance. The MAE of 3.15593255 is also moderate, better than the ARIMA model but worse than the LSTM. The R² value is slightly negative (-0.02979135), indicating that the ensemble model does not fit the data well, though it is not as poor as the ARIMA model. Compared to previous studies in [16,17], where CNN models often performed best for example Brazil's CNN model RMSE of 0.086, the ensemble model shows room for improvement, possibly by integrating more advanced deep learning techniques.

4.5 Overall Implications

The overall implication of this study lies in its contribution to the field of healthcare forecasting, particularly in the context of COVID-19 hospital admissions planning. By employing a rigorous methodology that integrates traditional time series analysis (ARIMA), deep learning (LSTM), and ensemble ARIMA-LSTM technique, we have provided insights into the predictive capabilities of these models in both public and private healthcare sectors. Our findings underscore the remarkable accuracy of LSTM models in forecasting hospital admissions, especially when compared to ARIMA and ensemble approaches. However, the ensemble model performs moderately but can potentially be enhanced by incorporating more sophisticated deep learning architectures or hyperparameter tuning. Additionally, our comparison against existing models highlights the generalisability and superior performance of LSTM models across different datasets and geographic locations, offering valuable insights for future research and practical applications in pandemic preparedness and management strategies.

5 Discussion and Conclusion

In this study, we explored the predictive capabilities of ARIMA, LSTM, and ensemble ARIMA-LSTM models in forecasting COVID-19 hospital admissions in both public and private healthcare sectors. Our results indicate that LSTM models consistently outperformed ARIMA and ensemble approaches in terms of accuracy and explanatory power. Specifically, LSTM models demonstrated highly accurate predictions with low RMSE and MAE values, as well as near-perfect R² values, suggesting an excellent fit to the actual data.

While the ensemble ARIMA-LSTM model showed improvement over the ARIMA model alone, it still fell short of the LSTM model's performance. This suggests that there is potential for further refinement of ensemble techniques, such as incorporating more advanced deep learning architectures or optimising hyperparameters [29].

The comparison against existing models from previous studies highlighted the generalisability of LSTM models across different datasets and geographic locations. Our LSTM model outperformed previous models in terms of accuracy and predictive power, indicating its robustness in forecasting COVID-19 hospital admissions.

In conclusion, our study contributes to the field of healthcare forecasting by providing insights into the effectiveness of various modelling techniques for predicting COVID-19 hospital admissions. LSTM models emerged as the most promising approach, offering accurate and reliable predictions. However, there is still room for improvement in ensemble techniques, and future research could explore more sophisticated deep learning architectures or incorporate additional features for enhanced prediction accuracy [30].

Overall, our findings have practical implications for pandemic preparedness and management strategies, enabling healthcare practitioners and policymakers to make informed decisions based on accurate forecasts of hospital admissions.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. WHO, I., Coronavirus disease (COVID-2019) situation reports. Situation report–152, 2020.
2. Jassat, W., et al., Increased mortality among individuals hospitalised with COVID-19 during the second wave in South Africa. MedRxiv, 2021: p. 2021.03. 09.21253184.

3. Naidu, T., The COVID-19 pandemic in South Africa. Psychological Trauma: Theory, Research, Practice, and Policy, 2020. 12(5): p. 559.
4. Solanki, G., et al., COVID-19 hospitalization and mortality and hospitalization-related utilization and expenditure: Analysis of a South African private health insured population. Plos one, 2022. 17(5): p. e0268025.
5. Burger, P. and E. Calitz, Covid-19, economic growth and South African fiscal policy. South African Journal of Economics, 2021. 89(1): p. 3-24.
6. Edoka, I., et al., Inpatient care costs of COVID-19 in South Africa's public healthcare system. International Journal of Health Policy and Management, 2022. 11(8): p. 1354.
7. Meyer-Rath, G., et al., The role of modeling and analytics in South African COVID-19 planning and budgeting. PLOS Global Public Health, 2023. 3(7): p. e0001063.
8. Kaushik, S., et al., AI in healthcare: time-series forecasting using statistical, neural, and ensemble architectures. Frontiers in big data, 2020. 3: p. 4.
9. Sah, S., et al., Forecasting COVID-19 pandemic using Prophet, ARIMA, and hybrid stacked LSTM-GRU models in India. Computational and Mathematical Methods in Medicine, 2022. 2022.
10. Sahai, A.K., et al., ARIMA modeling & forecasting of COVID-19 in top five affected countries. Diabetes & metabolic syndrome: clinical research & reviews, 2020. 14(5): p. 1419-1427.
11. Cloete, J., et al., Rapid rise in pediatric COVID-19 hospital admissionduring the early stages of the Omicron wave, Tshwane District, South Africa. medRxiv, 2021: p. 2021.12. 21.21268108.
12. Hyndman, R.J. and G. Athanasopoulos, Forecasting: principles and practice. 2018: OTexts.
13. Ismail Fawaz, H., et al., Deep learning for time series classification: a review. Data mining and knowledge discovery, 2019. 33(4): p. 917-963.
14. Nyasulu, J. and H. Pandya, The effects of coronavirus disease 2019 pandemic on the South African health system: A call to maintain essential health services. African Journal of Primary Health Care and Family Medicine, 2020. 12(1): p. 1-5.
15. Ma, R., Zheng, X., Wang, P., Liu, H. and Zhang, C., 2021. The prediction and analysis of COVID-19 epidemic trend by combining LSTM and Markov method. Scientific Reports, 11(1), p.17421.
16. Nabi, K.N., Tahmid, M.T., Rafi, A., Kader, M.E. and Haider, M.A., 2021. Forecasting COVID-19 cases: A comparative analysis between recurrent and convolutional neural networks. Results in Physics, 24, p.104137.
17. Ketu, S. and Mishra, P.K., 2022. RETRACTED ARTICLE: India perspective: CNN-LSTM hybrid deep learning model-based COVID-19 prediction and current status of medical resource availability. Soft Computing, 26(2), pp.645-664.
18. Shahid, F., Zameer, A. and Muneeb, M., 2020. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. Chaos, Solitons & Fractals, 140, p.110212
19. Valente, E., Roiati, M. and Pugliese, F., 2022. Forecasting the number of intensive care beds occupied by COVID-19 patients through the use of Recurrent Neural Networks, mobility habits and epidemic spread data. Statistical Journal of the IAOS, 38(2), pp.385-397
20. Devaraj, J., Elavarasan, R.M., Pugazhendhi, R., Shafiullah, G.M., Ganesan, S., Jeysree, A.K., Khan, I.A. and Hossain, E., 2021. Forecasting of COVID-19 cases using deep learning models: Is it reliable and practically significant?. Results in Physics, 21, p.103817.
21. Seo, H., Ahn, I., Gwon, H., Kang, H., Kim, Y., Choi, H., Kim, M., Han, J., Kee, G., Park, S. and Ko, S., 2024. Forecasting Hospital Room and Ward Occupancy Using Static and

Dynamic Information Concurrently: Retrospective Single-Center Cohort Study. JMIR Medical Informatics, 12, p.e53400.

22. Stellwagen, E. and L. Tashman, ARIMA: The models of Box and Jenkins. Foresight: The International Journal of Applied Forecasting, 2013(30): p. 28-33.

23. Gers, F.A., D. Eck, and J. Schmidhuber. Applying LSTM to time series predictable through time-window approaches. in International conference on artificial neural networks. 2001. Springer.

24. Gers, F.A., J. Schmidhuber, and F. Cummins, Learning to forget: Continual prediction with LSTM. Neural computation, 2000. 12(10): p. 2451-2471.
25. Deng, Y., H. Fan, and S. Wu, A hybrid ARIMA-LSTM model optimized by BP in the forecast of outpatient visits. Journal of Ambient Intelligence and Humanized Computing, 2020: p. 1-11.
26. Siami-Namini, S., N. Tavakoli, and A.S. Namin. A comparison of ARIMA and LSTM in forecasting time series. in 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA). 2018. IEEE.
27. Tsan, Y.-T., et al., The prediction of influenza-like illness and respiratory disease using LSTM and ARIMA. International Journal of Environmental Research and Public Health, 2022. 19(3): p. 1858.
28. Zhang, R., et al., Comparison of ARIMA and LSTM in forecasting the incidence of HFMD combined and uncombined with exogenous meteorological variables in Ningbo, China. International journal of environmental research and public health, 2021. 18(11): p. 6174.
29. Guo, A., et al., Predicting cardiovascular health trajectories in time-series electronic health records with LSTM models. BMC Medical Informatics and Decision Making, 2021. 21: p. 1-10.
30. Khatibi, T. and N. Karampour, Predicting the number of hospital admissions due to mental disorders from air pollutants and weather condition descriptors using a stacked ensemble of Deep Convolutional models and LSTM models (SEDCMLM). Journal of Cleaner Production, 2021. 280: p. 124410.

“System e Down”: Citizens’ Perceptions of the Failures of e-Government Systems in Botswana

Godwin Kaisara[1], Khulekani Yakobi[2], Clayton Peel[3] and Admire Mare[1]

[1] University of Johannesburg, Johannesburg, South Africa
[2] Mangosuthu University of Technology, Durban, South Africa
[3]Namibia University of Science and Technology, Windhoek, South Africa
gkaisara@gmail.com

Abstract. Globally, the importance of e-governance platforms in delivering services to citizens has significantly changed the way in which elected officials, public servants and ordinary people interact. However, in the African context, the digitisation of public services raised numerous unfulfilled expectations of seamless electronic service delivery among the public which often has to endure long queues, filling of countless forms and travelling long distances. Given that digitisation, datafication and platformisation processes are predominantly driven by platform companies from the Global North, most e-government platforms and systems are not easily relatable to the African context. Because of the inbuilt colonial matrix of power, African governments are too dependent on hardware and software from Asia and North Africa. This article relies on virtual ethnography and qualitative context analysis to assess user perceptions of e-government platforms and system failures in Botswana. Users’ perceptions and sentiments of e-government system failures were extracted from both government and corporate social media Facebook pages announcing e-government system failures in Botswana. Thematic analysis was adopted, resulting in the emergence of four themes, namely; encouraging law-breaking, loss of confidence in government, exasperation and security concerns. The results are discussed and implications put forward.

Keywords: Botswana, e-government, failure, Africa, user perceptions, social media, sentiment analysis.

1 Introduction

The exponential increase in Information and Communication Technologies (ICTs) has fundamentally transformed modern society [1], leading to a burgeoning body of literature on the contribution of ICTs to various developmental agendas. The delivery of public services in various countries has also transitioned with the adoption of various forms of e-governance, to an extent that has reformed the way governments operate, deliver services and interact with citizens [2]. Thus, governments have followed the corporate world’s investment in the potential of ICTs by digitizing many of their own operations so as to transform service provision [3], [4]. Proponents of e-government platforms and systems argue that they help governments attain better efficiency and improve governance, as well as enhance citizen participation, and improve relations with citizens [5]. Citizens who have grown accustomed to ICT-mediated services in the private sector now expect similar levels of e-services from the government [2], [6]. Etoundi and colleagues [7] state that ICT adoption in Africa is likely to lead to socio-economic growth, poverty reduction, and ultimately contribute to the reduction of wars. Consequently, the gold rush mentality around the implementation of e-government systems in developing countries has been recorded, albeit with minimal impact on addressing the needs of most citizens, rather than the relative few [8], [9].

A number of scholarly literature on various aspects of e-government in Africa have been published over the last two decades [2], [3], [10]. The growing importance attached to e-government implementation of these interfaces is evidenced by the growing number of sub-Saharan countries rolling out e-government implementation roadmaps [11]. In Southern Africa, a plethora of studies have attempted to document the status, challenges and prospects of e-government in countries such as Botswana [12], Namibia [10], [13], South Africa [2], Zambia [14], and Zimbabwe [15]. Some contemporary studies [16], [17] discuss e-government through the lens of the Fourth Industrial Revolution (4IR), arguably highlighting the continued importance of e-government in the midst of evolving technologies.

While e-government adoption in sub-Saharan Africa has been widespread, most of the initiatives have not met expectations [5]. Important in critically analyzing the efficacy of such interfaces is the quality of “interoperability” between service providers and users [18], especially when considered “from a technical, conceptual, and user interface point of view” [18, p. 237]. Furthermore, the findings of Daramola and Ayo [9] that African countries’ adoption of e-government platforms has not served the majority of their citizens, underscores the present need for research on user perspectives of e-government services in those countries. Citizen participation is one of the cornerstones of e-government [5], but few studies in the African context evaluate e-government services from a citizen’s viewpoint [19]. Arguably, failure to consider the perspectives of citizens violates the tenets of democracy, thus contradicting the very stated goal of e-governance enhancing democracy, among other benefits. Given the dearth of people’s perspectives on e-governance in the extant literature, this study seeks to gauge citizens’ perceptions of failing e-government systems in Botswana. As argued by Maclean and Titah [4], there is a dearth of empirical research on the negative impact or public value destroyed by e-government systems. Therefore, this paper may be viewed as an attempt to contribute to filling this vacuity in literature, by using user sentiments to gauge reactions to e-government systems failure. The paper is structured as follows: firstly, we provide a brief overview of e-government in Africa; secondly, we focus our discussion on e-government in Botswana; thirdly, we elucidate the role of social media in the role of e-government; then we proceed to discuss the methodology, followed by the results and conclusions of the study.

2 e-Government in Botswana

E-government arrived in Africa as an imported concept, based on imported designs and agendas [20], [21], and the subsequent studies in African e-government are underpinned by Global North theories [22]. African countries, as did most developing countries, implemented e-government initiatives with the support and guidance of Western institutions [23], [24]. In spite of the intellectual and financial support from the developed Global North, there are many of e-government initiatives which have fallen far short of the purported developmental, economic and poverty alleviation benefits [9]. A number of reasons have been advanced for the e-government failures in developing countries. However, many of these theories fall within the design-reality gap framework. Dada [25] summarized the various design-reality gaps as follows:

- *Hard-Soft Gaps*: the difference between the actual technology (hard) and the reality of the social context (people, culture, politics etc.) in which the system operates (soft).

- *Private-Public Gaps*: the difference between the private and public sectors means that a system that works in one sector often does not work in the other. This is due to incompatibilities between systems designed for the private sector and the reality of the public sector into which the system is transferred.

- *Country Context Gaps*: the differences that users experience between e-government systems in developed and developing countries, which arises from the gap between a system designed for first world countries, and the reality of a developing country into which the system is transferred.

In the words of Daramola [9], while the intentions of e-government are noble, "*[t]he reality, though, is that African countries' adoption of e-government platforms hasn't served the majority of their citizens. Services like e-taxation, e-payment and e-billing are useful for the middle class and richer people. But e-government initiatives that would support and cater to poorer people are sorely lacking".*

The Botswana government adopted its National Information and Communications Technology Policy, named *Maitlamo*, in 2007. This was preceded by an e-readiness assessment conducted in 2004. *Maitlamo* acts as a roadmap towards the transformation of the socio-economic, cultural and political status quo by effectively implementing ICTs in government service delivery [26]. The *Maitlamo* policy informs various sub policies such as the Botswana National e-Government Strategy 2011-2016, and the Botswana e-Government Master Plan 2015-2021. Recently, the Botswana government introduced the SmartBots initiative. The SmartBots initiative seeks to take advantage of the opportunities presented by the Fourth Industrial Revolution (4IR) and move the country towards the Knowledge-Based Economy. The SmartBots initiative has received donor funding from supranational entities such as the European Union, which funds the development of a National Innovation Capability Framework and a Digital

Business Package for Women Entrepreneurs [27]. The Botswana government aims to deliver a smart and efficient public sector through digitization of public services.

Schuppan [23] comments that sub-Saharan public services are punctuated by inefficiencies, limited capacity, and poorly trained personnel. As rightfully noted by Schuppan [23], e-government then becomes the remedial tool that could address some of the inefficiencies bedeviling the public sector. This is certainly true for Botswana, which has the stated goal of improving the convenience, quality and efficiency of service delivery [28]. However, the jury is still out on the extent to which e-government implementation has transformed public service delivery in Botswana. After a study assessing e-government in Botswana, Okike and Lobadi [12, p. 89] concluded that Botswana has "one of the leading progressive e-government systems on the African continent". Similarly, Bwalya and Healy [29] identified Botswana (along with South Africa, Mauritius and Seychelles) as one of the leaders in e-government in southern Africa.

As observed by the researchers, anecdotal evidence from various government announcements on social media paint a picture of e-government systems that are failing. The phrase "System e down" has become a frequent headline and catchphrase among e-government commentators and users in Botswana. While Okike and Lobadi [12] established that Botswana citizens were satisfied with e-government services rendered, commentary from social media suggests the contrary.

3 Usage of social media as part of e-government systems and platforms

Social media has become an indispensable source of information in the knowledge society, playing an important role in defining citizen participation and action, through discussions and coordination of activities [30], [31]. Stieglitz and Dang-Xuan [32] state that the growth of social media has made it an invaluable tool to gauge citizens' opinions and attitudes on particular issues, including their perceptions of government services provision. In Botswana, the government has been using social media platforms since 2011, although there are no documents guiding the government's social media use [33]. It has thus become prudent for policymakers to analyze social media data for a reflection of citizens' sentiments [18]. Consequently, Social Media Analytics (SMA) has become increasingly critical for institutions and governments [32], although SMA literature in the context of e-government is still very limited [31].

4 RESEARCH METHODOLOGY

4.1 Research philosophy and design

In order to identify how Botswana citizens perceive the failed e-government systems, we adopted a qualitative design, underpinned by an interpretive research philosophy. The interpretivist philosophy situates research in a particular context, and views reality as a social construct that may not necessarily be universal. Furthermore, scholars such as Haro-de-Rosario and colleagues [28, p. 44] have called for more qualitative studies of citizen/e-government interaction, stating that "it would also be interesting to conduct a more qualitative, rather than quantitative, analysis of citizens' comments and responses to the government…". Hence our adoption of a qualitative stance may be regarded as a response to such calls from the scholarly community. The study employed a virtual ethnography method to watch how citizens reacted to Facebook posts about failing Botswana e-government systems. Hine [29] states that highlighting the multifaceted nature and social structure of social media-based interaction is the aim of bringing ethnography to online settings.

4.2 Data collection

The analyses period for the government posts, the reactions and comments from citizens, was between 2011 to 2023. The comments from users were extracted from both government and corporate social media Facebook pages and saved to a Google Drive spreadsheet using the Groupboss application, no filtering was applied. By going to each citizen's Facebook profile page who left a comment, it was simple to determine the location/nationality of the individual based on that profile. Our search on Facebook brought up many pages on "system e down", with each post having numerous comments about the country's e-government systems failing.

4.3 Data analysis

We used thematic analysis to examine citizens' reactions to any announcements relating to Government Information System (un)availability posted on the official Botswana government Facebook page, using Facebook comments as a unit of analysis. We used Facebook as a testbed for two reasons. Firstly, studies have revealed that Facebook is a preferred medium when participating in government matters [30]. Secondly, Facebook has a much wider footprint than its competitors in Botswana [34]. Comments were coded at sentence level. Sentence level coding was chosen owing to the relatively short comments that individuals made on Facebook.

4.4 Trustworthiness and confirmability

Different approaches were employed to enhance trustworthiness of the study and confirmability. To enhance trustworthiness, the researchers judiciously used verbatim quotes to keep the analysis as close to the actual comments as possible. As argued by some scholars, “presenting original voices of participants can also enhance the rigour of qualitative research” [35, p. 169]. For confirmability, the researchers utilized peer debriefing. A colleague who was not involved with the research was asked to assess the emergent themes and associated quotes and give their feedback.

5 FINDINGS

Our search encountered several posts related to failing e-government systems such as the Government Integrated Financial Management Information System, University of Botswana, and the Department of Road Transport and Safety. Some of the posts are from nine years ago, thus indicating that consumer complaints about the unreliability of government online platforms is a persistent problem. Our analysis led to the emergence of four thematic areas, which are expounded upon in the following sub-sections.

5.1 Theme one: Encouraging law-breaking

Several posts suggested that some citizens may resort to committing illegal activities due to the e-government systems failure. When commenting on the transport e-government system, one commentator stated that “*even the transport system can take the whole week [off], whilst we are suffering…*”. The continued unavailability of e-government systems provokes thoughts of engaging in illegal activities. This is evidenced by one commentator who stated that “*…I will start driving ka [with an] ID. [I] am tired of all this honestly*”. In a related post, another citizen stated “*and then you want to act surprised when people buy licences?*”. Many of the posts indicated that because of the regular and often long periods of system unavailability, citizens had to drive without a license. Others added that they had been fined by the police for driving with expired licenses. All this perhaps explains why some citizens had to resort to irregular measures to get a licence, be it authentic or fake.

5.2 Theme two: Loss of confidence

Various comments showed that there was an increasing loss of confidence in the government’s ability to deliver services through e-government. In response to a government statement announcing the unavailability of the e-government services at the department of Road Transport and Safety, a number of less than complementary comments were made. For example, one commentator labeled it *“the most hopeless and* *useless department"*. Another labeled it *"the worst department and ministry in bw"* (identifying the country by its domain) and added that the *"minister is also clueless.",* while yet another user simply said, *"You people are really useless"*. While the foregoing comments may imply annoyance, other comments suggested a dismissive attitude due to the recurrence of the problem. For example, one user argued that it is *"rather permanent, 'network e down' is their daily response."* Sharing similar sentiments, one user said that *"another name for Botswana is SYSTEM E DOWN. This has been going on for years"*. As result, some comments suggested that the government consider reverting back to manual systems to deliver services to citizens, as government was not ready for effective e-government service delivery. Comments questioning Botswana's readiness for e-government include the following:

"Let's go back to pen and paper, this system thing delays payments".
"A big NO NO NO, Botswana is not ready".
"Manually, please".
"Things of this place…we are still in the 1960s".

The perpetual unavailability of e-government systems led some citizens to question if there were funds and capacity to effectively implement and manage e-government systems. The loss of confidence also extended to the leadership in government. Comments alluded to the political leadership being both incompetent and self-serving, as evidenced by one user who commented that "*hopeless in all ways, their systems work only a day in the entire year, the minister is too worried about his personal agenda, rather than getting this nonsense sorted".*

5.3 Theme three: Exasperation

Some comments suggest that some users have become apathetic and accepting of the persistent system failures as a reality. For example, one commentator stated that they had "*learnt 2 acept "system e down" as our 2nd national anthem n d atitude thrown at u by d persn hu says dose words*". Another, responding to a post about system failure at the University of Botswana simply commented "*who the hell cares*", while another also "*what else is new*".

5.4 Theme four: Security concerns

Concerns were also raised about the security implications of failing e-government systems. In one post lamenting the prevalence of e-government system failures, one user stated the following:

"…system e down at the border gates is going to cost us big time, [because it means that] criminals cross the border without any detection or trace.......if they can't even notice Kgosi the former DISS Director, how will they notice other criminals? We should

be worried as a country, this is a risk to our National Security. [It implies that] those wanted by the police could have easily sneaked out of the country or into the country running away from other countries".

Responding to the initial comment raising security concerns, one of the users supported the view, stating; "*I know that you speak from a political perspective but, still very true from a general point of view*".

6 Discussion of findings

The comments of Botswana citizens on Facebook regarding e-government unavailability revealed varying negative reactions to the ineffectiveness of the e-government system. Nevertheless, the reactions suggest that Botswana, and other developing countries, need to be intentional in their efforts to arrest the well-documented e-government failures. The theoretical and practical implications are discussed next.

6.1 Theoretical implications

Although the findings indicate that e-government failure may encourage law-breaking, to the best of our knowledge, no studies have highlighted the link between e-government failure and increased intentions to break the law. While literature indicates that e-government may be used to enhance law enforcement [36], the findings also indicate that poor-performing e-government systems may encourage anti-social behaviors .

The findings reveal that e-government has a deleterious effect on citizenry's confidence in government. Not only do e-government failures lead to a loss of confidence in government, but Twizeyimana and Andersson [37] add that it leads to resistance to any future e-government initiatives. Similarly, a study conducted in Liberia by Mensah and colleagues [38] established that poor e-government performance does have a deleterious effect on citizens' confidence in government. Furthermore, this is counter to one of the goals of e-government implementation, which is to improve citizens' confidence in their governments. The loss of confidence has negative implications on the government's legitimacy, as government's legitimacy is closely tied to citizens' confidence in its ability to perform [39].

Our findings suggest that regular e-government failure may also cause exasperation among users. The findings are consistent with the assertion of Twizeyimana and Andersson [37], that e-government failure may lead to distress. Furthermore, exasperation may lead to a decrease in motivational levels to use e-government services. Arguably, this may also contribute towards strained relations between citizens and government.

Implications of e-government failure on security provision were also raised, particularly the potential failure to manage and monitor the movement of people. The integral role of technology in immigration control, particularly monitoring and identifying people on the watch list, is highlighted by Maulana [36]. However, most of the extant e-government literature focuses on security in online settings [2], [38]. Relatively little has been written about the offline security implications of e-government systems failure. Although online security is inarguably important, particularly where there is an exchange of information, e-government failure has offline security implications that policymakers need to be cognizant of.

6.2 Practical implications

The study highlights the importance of reducing e-government failure rates, which are still prevalent in the developing world. Since the results indicate that e-government failure leads to some unintended consequences or undesirable behavioral patterns, e-government practitioners need to prioritize the development of more efficient e-government systems to minimize the aforementioned consequences. E-government failure should not only be viewed as a resource drain, but contributing to the emergence of some undesired attitudes and behaviors.

7 Conclusion

E-government failure in the developing world is well documented. On the other hand, relatively few studies have sought to understand how e-government users in developing countries react or perceive such occurrences. Consequently, this paper sought to shed light on how citizens of a developing country react to perpetual e-government failure. Comments of disgruntlement on the government's official Facebook page content were successfully collected and analyzed. The trends of the comments highlighted a key phrase of "system e down" which significantly highlighted the failure of the e-government system in Botswana. The limitations of the paper are that only one social network was used to analyzed citizens' comments. The analysis was limited to the written comments and sentiments, as the study did not analyzed images and emojis. There were also limitations associated with the use using virtual ethnography methodology. Virtual ethnography allowed us to collect comments of active Facebook users, meaning that the results or sentiments could not be generalized to the broader population. Therefore, the views of non-Facebook users did not make part of the study. Online interactions and analysis of online textual sentiments may lack the richness of non-verbal cues present in face-to-face interactions. Future recommendations are that scholars need to adopt methodologies that will allow face to face interactions and behaviors from citizens. Researchers can also adopt explanatory research approach to test hypothesis propositions that may emerge from citizen reactions. Scholars and practitioners should also adopt social media analytics methods and tools in the context of e-government to monitor social media data for the purposes of analyzing data that reflect citizens' concerns and needs. Social media analytics may offer a range of tools and methods to monitor, analyze and visualize big data from various social media platforms to generate insightful contents and analytics.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] H. Wimelius, L. Mathiassen, J. Holmström, and M. Keil, "A paradoxical perspective on technology renewal in digital transformation," *Inf. Syst. J.*, vol. 31, no. 1, pp. 198–225, 2021, doi: 10.1111/isj.12307.

[2] G. Kaisara and S. Pather, "The e-Government evaluation challenge: A South African Batho Pele-aligned service quality approach," *Gov. Inf. Q.*, vol. 28, no. 2, 2011, doi: 10.1016/j.giq.2010.07.008.

[3] K. J. Bwalya, "Factors Affecting Adoption of e-Government in Zambia," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 38, no. 1, pp. 1–13, 2009, doi: 10.1002/j.1681-4835.2009.tb00267.x.

[4] D. MacLean and R. Titah, "A Systematic Literature Review of Empirical Research on the Impacts of e-Government: A Public Value Perspective," *Public Adm. Rev.*, vol. 82, no. 1, pp. 23–38, 2022, doi: 10.1111/puar.13413.

[5] Q. N. Nkohkwo and M. S. Islam, "Challenges to the successful implementation of e-government initiatives in sub-Saharan Africa: A literature review," *Electron. J. e-Government*, vol. 11, no. 2, pp. 253–267, 2013.

[6] J. Chipeta, "A review of e-government development in Africa: A case of Zambia," *J. e-Government Stud. Best Pract.*, pp. 1–13, 2018, doi: 10.5171/2018.973845.

[7] R. A. Etoundi, S. F. M. Onana, A. A. Eteme, and M. L. F. Ndjodo, "Special issue on ict for Africa development: An introduction and framework for research," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 76, no. 1, pp. 1–11, 2016, doi: 10.1002/j.1681-4835.2016.tb00551.x.

[8] O. Daramola and C. Ayo, "Enabling socio-economic development of the masses through e-government in developing countries," in *European Conference on Digital Government*, 2015, pp. 508–516.

[9] O. Daramola, "African countries should rethink how they use e-government platforms," *The Conversation*, 2019. https://theconversation.com/african-countries-should-rethink-how-they-use-e-government-platforms-108689 (accessed Apr. 02, 2024).

[10] K. Amukugo and A. Peters, "Citizen-centric e-government services in Namibia: Myth or reality?," in *First African Conference on Human Computer Interaction*, 2016, pp. 193–197. doi: 10.1145/2998581.2998610.

[11] S. F. Verkijika and L. De Wet, "E-government adoption in sub-Saharan Africa," *Electron. Commer. Res. Appl.*, vol. 30, pp. 83–93, 2018, doi: 10.1016/j.elerap.2018.05.012.

[12] E. U. Okike and N. O. Lobadi, "An assessment of e-government programme in Botswana," *Int. J. Comput. Sci. Inf. Secur.*, vol. 17, no. 12, pp. 81–91, 2019.

[13] C. T. Nengomasha and T. N. Shuumbili, "Access to e-government services by

citizens through public/community libraries in Namibia," *Inf. Dev.*, vol. 38, no. 1, pp. 68–82, 2022, doi: 10.1177/0266666920979009.

[14] Z. Chilembo and S. Tembo, "Opportunities and challenges of coordinating the implementation of e-government programmes in Zambia," *Int. J. Inf. Sci.*, vol. 10, no. 1, pp. 29–43, 2020, doi: 10.5923/j.ijis.20201001.04.

[15] E. Ruhode, "E-Government for development: A thematic analysis of Zimbabwe's information and communication technology policy documents," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 73, no. 1, pp. 1–15, 2016, doi: 10.1002/j.1681-4835.2016.tb00532.x.

[16] G. Kaisara, A. Mare, and C. Peel, "Haven't we been here before? A critical analysis of the fourth industrial revolution," in *2021 Conference on Information Communications Technology and Society, ICTAS 2021 - Proceedings*, 2021, no. May, pp. 67–72. doi: 10.1109/ICTAS50802.2021.9395028.

[17] S. Layton-Matthews and C. Landsberg, "The Fourth Industrial Revolution (4IR) and its Effects on Public Service Delivery in South Africa," *Think.*, vol. 90, no. 1, pp. 55–64, 2022, doi: 10.36615/thethinker.v90i1.1173.

[18] A. Androutsopoulou, N. Karacapilidis, E. Loukis, and Y. Charalabidis, "Towards an integrated and inclusive platform for open innovation in the public sector," in *International Conference on e-Democracy*, 2017, vol. 792, no. November, pp. 228–243. doi: 10.1007/978-3-319-71117-1_16.

[19] K. Fröhlich and A. Peters, "E-government social exclusion and satisfaction among namibian citizens: A case of a namibian government ministry," *ACM Int. Conf. Proceeding Ser.*, vol. Part F1308, 2017, doi: 10.1145/3129416.3129435.

[20] R. Heeks, "e-Government in Africa: Promise and Practice," 2002. doi: 10.1515/9781626373839-008.

[21] T. Mawela, N. M. Ochara, and H. Twinomurinzi, "E-government implementation: A reflection on South African municipalities," *South African Comput. J.*, vol. 29, no. 1, pp. 147–171, 2017, doi: 10.18489/sacj.v29i1.444.

[22] B. Mukabeta Maumbe, V. Owei, and H. Alexander, "Questioning the pace and pathway of e-government development in Africa: A case study of South Africa's Cape Gateway project," *Gov. Inf. Q.*, vol. 25, no. 4, pp. 757–777, 2008, doi: 10.1016/j.giq.2007.08.007.

[23] T. Schuppan, "E-Government in developing countries: Experiences from sub-Saharan Africa," *Gov. Inf. Q.*, vol. 26, no. 1, pp. 118–127, 2009, doi: 10.1016/j.giq.2008.01.006.

[24] V. Ndou, "e-Government for developing countries: opportunities and challenges," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 18, no. 1, pp. 1–24, 2004.

[25] D. Dada, "The Failure of E-Government in Developing Countries: A Literature Review," *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 26, no. 1, pp. 1–10, 2006, doi: 10.1002/j.1681-4835.2006.tb00176.x.

[26] M. Bakwena and Z. Kahaka, "The Botswana national Information and Communication Technology policy and economic diversification: How have e fared thus far?," *Botsw. Notes Rec.*, vol. 45, pp. 206–213, 2012, [Online]. Available: http://www.ub.bw/ojs/index.php/bnr/article/viewFile/411/170

[27] A. Maramwidze, “EU confirms support for Botswana’s SmartBots strategy,” *IT Web*, 2021. https://itweb.africa/content/Pero3qZx9ad7Qb6m (accessed Apr. 03, 2024).

[28] C. B. Mokone, O. T. Eyitayo, and A. Masizana- Katongo, “Critical success factors for e-government projects: The case of Botswana,” *J. e-Government Stud. Best Pract.*, vol. 2018, pp. 1–14, 2018, doi: 10.5171/2018.335906.

[29] K. J. Bwalya and M. Healy, “Harnessing e-Government Adoption in the SADC Region : a Conceptual Underpinning,” *Electron. J. E-Government*, vol. 8, no. 1, pp. 23–32, 2010.

[30] A. Haro-de-Rosario, A. Sáez-Martín, and M. del C. Caba-Pérez, “Using social media to enhance citizen engagement with local government: Twitter or Facebook?,” *New Media Soc.*, vol. 20, no. 1, 2018, doi: https://doi.org/10.1177/146144481664565.

[31] K. Yakobi, B. Scholtz, and B. W. vom Berg, “Benefits and applications of social media analytics for citizen relationship management,” in *19th European Conference on Digital Government*, 2019, pp. 178–187.

[32] S. Stieglitz and L. Dang-Xuan, “Social media and political communication: a social media analytics framework,” *Soc. Netw. Anal. Min.*, vol. 3, no. 4, pp. 1277–1291, 2013, doi: 10.1007/s13278-012-0079-3.

[33] T. L. Mosweu, “The use of social media by the Botswana government and governance of liquid communication,” *Mousaion South African J. Inf. Stud.*, vol. 37, no. 1, 2019, doi: 10.25159/2663-659x/5967.

[34] Statista, “Number of social media users in Botswana as of January 2024, by platform (in 1,000s),” *Statista*, 2024. https://www.statista.com/statistics/1303916/most-used-social-media-in-botswana/ (accessed May 20, 2024).

[35] A. Younas, S. Fàbregues, A. Durante, and P. Ali, “Providing English and native language quotes in qualitative research: A call to action,” *Nurs. Open*, vol. 9, no. 1, pp. 168–174, 2022, doi: 10.1002/nop2.1115.

[36] R. N. Maulana, “The Indonesian e-government initiatives: Development stages of e-government in immigration law enforcement,” *J. Ilm. Kaji. Keimigrasian*, vol. 4, no. 2, pp. 81–103, 2021, doi: 10.52617/jikk.v5i2.263.

[37] J. D. Twizeyimana and A. Andersson, “The public value of e-government - a literature review,” *Gov. Inf. Q.*, vol. 36, pp. 167–178, 2019.

[38] R. Mensah, A. Cater-Steel, and M. Toleman, “Factors affecting e-government adoption in Liberia: A practitioner perspective,” *Electron. J. Inf. Syst. Dev. Ctries.*, vol. 87, no. 3, pp. 1–15, 2021, doi: 10.1002/isd2.12161.

[39] M. Kazemi, M. Gharibi, H. M. Mosammam, and M. F. Ghal’e, “The effect of e-government service quality on public trust: Case study: Saanat o Madan Bank of Iran,” *Bankac*, vol. 2, no. 7.8, pp. 4–18, 2015, doi: 10.1501/bsad_0000000017.

Pointers for Ubuntu Information Systems Ethics

Jan H. Kroeze[1][0000-0001-7118-4853]

[1]University of South Africa (Science Campus), Roodepoort, South Africa
kroezjh@unisa.ac.za

Abstract. Although there has been interest in the ethical aspects of information systems (IS) since the 1980s, various authors have recently lamented the fact that not enough research has been done in the area and that ethics is often ignored in the IS industry. When one searches for research on the decolonisation and Africanisation of IS ethics, few outputs can be found. The main research question that this article addresses is: How can African knowledge systems and ways of knowing inform and enrich IS ethics? The main aim of the article is to identify appropriate ethical insights, borrowed from Ubuntu-informed ethics, information ethics and business ethics to serve as pointers for Ubuntu IS ethics. The study is a conceptual study which follows a philosophical approach. The research is a rudimentary attempt to enrich IS ethical theory from an African viewpoint. The most important contribution of the paper is the proposal of a root for an Ubuntu-based IS ethic which could be used to counteract the hegemony of Eurocentric values embedded in information and communication technology.

Keywords: Ubuntu, Information Systems (IS), Ethics, Diversity, Africanisation, Decolonisation.

1 Introduction

Although there has been interest in the ethical aspects of information systems (IS) since the 1980s, various authors have recently lamented the fact that not enough research has been done in the area and that ethics is often ignored in the IS industry [1–3]. When one searches for research on the decolonisation and Africanisation of IS ethics, few outputs can be found. While some relevant material can be found in closely related fields, such as African ethics, African information ethics and African business ethics, no one has – to my knowledge – attempted to integrate and apply the bits and pieces to the IS field.

The research question that this article addresses is: How can African knowledge systems and ways of knowing inform and enrich IS ethics? Already in the previous century, Samuel [4] noted that Africans represent a rich diversity of cultures and values. One should, therefore, be careful not to generalise ethical principles from one African culture, assuming that it is valid for the whole continent. This paper focuses on Ubuntu-informed value systems as one possible permutation of African ethics. Ubuntu is a holistic but complex system of thought that spans how people understand the physical world and the essence of being human, as well as human experiences, values and lifestyles [5]. Otaluka regards Ubuntu as “an ancient moral theory” [6]. Ubuntu and related value systems are typical of sub-Saharan Africa (SSA) [7]. The paper answers the research question by indicating:

- How the theoretical principles of global ethics, information ethics and business ethics inform IS ethics
- How the theoretical principles of Ubuntu, Ubuntu ethics, Ubuntu information ethics, and Ubuntu business ethics amend IS ethics
- How the core of an Ubuntu IS ethics can be formulated to guide the design and use of IS in communalist environments

The paper contributes to the SAICSIT 2024 conference theme of Human-Machine-Digital Convergence with specific reference to the track on IS for sustainable development and the United Nations’ sustainable development goal to “[r]educe inequality within and among countries” (https://sdgs.un.org/goals/goal10). Not only does it create awareness that Western value systems are often embedded in IS, but it also provides an ethical foundation for the adoption, adaptation and creation of IT artefacts within communalist contexts. The discussion centres its attention on the ethical issues around the design and impact of IS, but a few comments will be made in passing about IS theory and research because it grounds practice.

2 Global Ethics

This section positions IS ethics within the broader field of global ethics because it is important to understand the general concepts of information ethics and business ethics before an Ubuntu perspective of these concepts can be offered. Ethics as an academic discipline philosophises on the nature and relevance of value systems in all scientific fields [2]. Kantian deontology and utilitarianism are two of the most important streams in Western ethics. Deontology strives to formulate a single ethical principle used to determine if an action is ethical or unethical, while utilitarianism evaluates the effects of an action to judge whether it is good or bad. According to Lajul [8], both these approaches are reductionist and based on individualistic values.

Information ethics is an important reference discipline for IS because IS deals with the electronic storage, processing and use of information in the business industry, organisations and society, as well as the effect of digital applications on society and the environment. Information ethics also deals with ethical issues related to information and communication technology (ICT) and social media [9]. Durani, Eckhardt and Kollmer [10] regard business ethics as a higher level of normative ethics where meta-ethical considerations are integrated with applied ethical theories. IS ethics is closely related to business ethics but may have a wider and more philosophical scope. Western ethical principles have been standardised in mainstream, globalised IS ethics [11]. Eurocentric worldviews are deeply embedded and dominant in IS research and theory [12]. As an alternative principal scope, African ethics may provide a valuable lens to enrich global IS ethics and to guide the actuation of IS ethics on the continent.

IS scholars and practitioners should realise that “computer technologies are not neutral – they are laden with human, cultural and social values” [9], for example in artificial intelligence (AI) [13]. The IS guild should, therefore, acknowledge its ethical responsibilities when designing, building, implementing and evaluating software [13]. Since ICT artefacts can often have dual properties and can be used to change the world positively or negatively [10], decision-makers in the ICT sphere must be aware of their moral duties and should therefore be trained to deal with potential ethical problems. Where many stakeholders are involved, the locus of responsibility is problematic: who is responsible for software that is intended to prompt unethical behaviour or for the unintended, possibly harmful use of applications – the system owners, designers, programmers or users [14]? The Giving Voice to Values framework may be a helpful tool to provide ethical training and dialogue skills to system owners, designers and users to communicate these issues constructively [15].

Ethics is especially important in certain subfields of IS, such as decision support systems (DSS) and AI, where the software does not merely automate basic operational systems but is knowledge-based and automatically makes judgements to provide normative support. System designers and ICT practitioners have a moral duty "to highlight potential ethical issues in proposed systems" [14]. With limited research available regarding the ethics of DSS, Meredith and Arnott [14] suggest that the bioethical principles of beneficence, non-maleficence, autonomy and justice should be applied to the field of DSS to guide ethical decision-making because there are many similarities between clinical and general decision-making and -support [8].

Design is one of the most salient aspects of IS. "[T]he ethics of design refers to challenging the status quo and can be understood as disclosing and exposing the hidden nature and motivations, ideologies, and interests embedded in DSR [Design Science Research]" [10]. Unfortunately, ethical issues around design have not yet received sufficient attention [3, 10].

Social media contain embedded capabilities, driven by AI algorithms, which could be used unknowingly or on purpose in unethical ways harming the community or society [16]. The harnessing of user profiles by social media to filter content to please users may have unintended adverse consequences. Filter bubbles (curating content based on a user's historical preferences) narrow individuals' worldviews by restricting their access to balanced information [16, 17]. AI is often used to generate these suggestions for further reading to social media users [16].

Moreover, AI has recently received a lot of attention due to the arrival of conversational AI applications such as ChatGPT and Bing. These intelligent chatbots communicate with human users by providing confident answers to their questions in fluent human language creating the impression that all answers are correct and trustworthy. This creates an array of ethical issues, not only because the answers are sometimes inaccurate, but also regarding the copyright of the sources used for the compilation of the answers. Other ethical issues that are related to AI pertain to data surveillance [17], consumerisation caused by social media [18], sexual and reproductive technologies [19–22], and the negative implications for job creation [23]. According to Prabhakaran, Mitchell, Gebru, and Gabriel [13], AI models can be "biased, unfair, or unethical". Prabhakaran et al. [13] suggest that human rights should be used as a basis to guide the ethical design and use of AI because these values are accepted and used globally.

However, they admit that it is just a starting point which should be amended with local cultural values to avoid the danger of neo-imperialism.

With the landscape of global ethics, information ethics, business ethics and IS ethics outlined, the essay can move the looking glass to explore associated aspects in the context of SSA.

3 Ubuntu Ethics

This section focuses on Ubuntu ethics, information ethics and business ethics which provide relevant pointers for Ubuntu IS ethics. African ethics offer alternative principles to amend IS theory and practice since a value-based information system "incorporates human and ethical values into its core design" [3]. Except for the Ubuntu principle that has been used for some attempts to work out an African scholarly philosophy – compare, for example, Metz's [7] nascent African moral theory – detailed, systematic African epistemologies and ethical systems do not yet exist. "Africa does not seem to have any clearly *documented* ethical and moral traditions [emphasis added]" [24].

Ubuntu can contribute to post-positivist thinking by decentring empiricist thinking. It could enrich business ethics by providing a lens to decontextualise Western knowledge [19]. It also provides "an ethic of inclusiveness" as a constructive principle to overcome the divisive heritage of colonialism [25]. Papers by Chirongoma and Mutsvedu [26], Sande [27] and Moyo [20] are only three recent examples of scholarly research that use Ubuntu to provide an African perspective on a variety of ethical issues in the digital era, such as the impact of ICT on human relationships, religion and even moral issues concerning the use of sex robots.

3.1 Basic Ubuntu Principles

Okyere-Manu [19] identifies the following Ubuntu ethical principles: "Characteristics associated with Ubuntu in the literature include having harmonious relations with others, belongingness, sharing identity with the community and showing solidarity with others." Ubuntu axiology prioritises human dignity and relationality over money and economy [5]. Based on the unique African view on ethics that moral values are founded in human relationships, Metz [7] formulates the following basic ethical principle as the foundation for a formalised African-Ubuntu ethic: "An action is right just insofar as it promotes shared identity among people grounded on good-will; an act is wrong to the extent that it fails to do so and tends to encourage the opposites of division and ill-will." Sande [27] provides a concise version of the same principle: "an action is right to the extent that it maximizes harmony".

Naude [28] believes that the dichotomy of individualism versus communalism to differentiate Western and African ethics is false. According to Naudé [28], the tendency to live in cooperative communities and to conform to group norms ("sociality") is also an important Western value, while an individual's uniqueness and responsibility ("personhood and autonomy" are also inherent in African cultures [25]. Lajul [8] agrees: "In as much as the community is important in most African cultures, they equally honour and respect individual differences." It may be true, however, that there is a difference in emphasis. Western ethics focuses on the individual, the organisation and society at large but is relatively quiet about smaller, more immediate communities, while Ubuntu ethics acknowledges the role of individuals in their more immediate communities but is quieter about society at large [29]. While an Ubuntu worldview also acknowledges the human person, it sees it as a "corporate individual" which means "that human beings are individuals who are autonomous, free and self-propelling but at the same time, they are ontologically related to others, nature and the spiritual world" [8]. This could explain why African philosophers use the dichotomy as a foundation to construct an alternative to Western ethics [28]. This leaves room for an African perspective to enrich and complement IS ethics.

When considering Ubuntu ethics as a source of enrichment for IS ethics, scholars should do this in a critical way, because constructive critique is the essence of scientific thinking. While the communalist perspective could be an important contribution towards ethics, it may also stimulate new critical questions. Meredith and Arnott [14] emphasise the locus of responsibility as residing within a person to regard behaviour as right or wrong. If there is no personal ownership, acts cannot be regarded as either moral or immoral and bad things are simply accidents. Communalism as a point of departure in ethics may become problematic in some African cultures if the community is regarded as more important than the individual. Who takes responsibility for immoral behaviour if an individual in a community cannot be held accountable?

A balance between the two values is therefore important. Too much emphasis on the community may lead to factionalism and nepotism [6], or a lack of ownership and responsibility which stifles innovative technologies created by Africans to address their own needs [30]. Too much emphasis on the individual could again lead to selfishness, weak social networks and deficient mutual support, which could in turn be problematic in some Western cultures. One should add that there is nothing wrong with reinterpreting and adapting traditional values for a new era; for example, the scope of Ubuntu can be broadened from the local tribe to all humanity, and the scope of Western ethics can be amended by allocating equal importance to communalist values compared to the value of the individual and the society at large.

3.2 Ubuntu Information Ethics and Business Ethics

Ubuntu information ethics is an important field that could contribute to Ubuntu IS ethics. Britz [31] discusses several factors that lead to information poverty in Africa which could again hamper development and growth. Diverse ontologies and cultural backgrounds may result in an experience of alienation. The dominance of English as the lingua franca of the digital world creates tough barriers for non-English speaking communities. The role of gender and status in communities also impacts the way information is interpreted. There is a need for IS research to address these factors to overcome the issue of information poverty. Custom-made applications that translate information into idiomatic local languages and explain cultural differences could go a long way to facilitate solutions in this regard.

Capurro [9] mentions Ubuntu as a unique African philosophy that can contribute to information ethics by reflecting on how ICTs affect African communities. Many Africans in former colonies live in two worlds, i.e., a Western lifestyle in urban environments versus an Ubuntu lifestyle in rural communities, or a combination of the two. There is a need for an African-to-Western flow of information to balance the stereotypic, opposite drift to ensure reciprocal respect for the different cultures. This mutual understanding of and respect for diverse cultures is important since culture influences moral values [32]. While it is important to identify and develop a unique Ubuntu-infused African information ethic, this should not be done in isolation but in interaction with the rest of the world so that the various systems can complement and enrich each other – "to develop ethical Information Ethics models that have wider applicability and validity beyond nations, regional and continental boundaries" [24].

Since IS ethics is closely related to business ethics, African-Ubuntu business ethics should also be used to amend its meta-ethical foundations. In a certain sense, one should ask whether it is necessary to Africanise business ethics because African scholars' thinking made significant contributions toward Western philosophy [28]. If one accepts that Western philosophy currently dominates contemporary axiology (the theory of values), these Eurocentric values could be 'translated' into African contexts by illuminating the principles using local perspectives, using African case studies to explain Western ideas, or solving African moral issues using Western insights. It is rather obvious that both direct transfer and translation make only small contributions to decolonise axiology [28]. Constructing a unique, African business axiology is needed to make a substantial contribution.

An example of a unique African business ethic can be found in Taylor's work [33]. Based on Metz's basic Ubuntu ethics principle referred to above, Taylor [33] proposes the following Ubuntu-based business ethical principle: "An action is right insofar as it promotes cohesion and reciprocal value amongst people. An action is wrong insofar as it damages relationships and devalues any individual or group." Business actions are tested against the four heuristics of the business ethic [33]:

- "Does the action promote cohesion amongst the parties?"
- "Does the action promote or acknowledge reciprocal value between the parties?"
- "Does the action damage relationships with the various parties?"
- "Does the action devalue any of the parties?"

Woermann and Engelbrecht [34] propose a relation-holder ethic rather than a stakeholder ethic as the basic foundational principle for business ethics. Woermann and Engelbrecht [34] reinterpret the principle of harmony and a shared identity to the business domain as follows: "The firm has a duty to foster harmonious interpersonal relations with (potential) parties who have the capacity to: a. Identify with, and show solidarity towards, the firm; b. Be identified with, and to benefit from, the firm's solidarity." The principle implies that not only the management team of a business but also its employers and the affected community should participate in strategic decision-making.

Constructive attempts to formulate Ubuntu-informed ethical principles and to apply these in business ethics, such as those presented above, provide directions for the development of Ubuntu IS ethics.

4 Ubuntu IS Ethics

This section reflects on existing research that deals with ethical issues regarding computer hardware, programming languages, and software applications. The need for an Ubuntu IS ethic is identified to mitigate these issues. A few pointers are provided as possible foundations for such an ethics. A basic ethics nucleus is proposed with support from recent literature. Some implications for IS scholars and practitioners are also suggested.

Lamola believes that Western norms are built deeply into the technologies that drive the fourth industrial revolution [35], which represents a recent embodiment of the epistemic domination of information technology [36]. According to Prabhakaran et al. [13], "the Western values implicitly encoded into AI systems may be at odds with other value systems, creating the risk of problematic value imposition when these technologies are deployed globally." Users are, of course, free to adopt or reject any software system, but in practice, there are often no alternatives, thus leaving them with little or no choice other than using what is available. The field should look for ways to create room for alternative (even contradictory) views on reality and knowledge to co-exist and be valued equally [37]. This prompts the need for African IS ethics to guide the ethical design and use of IS on the continent.

Ubuntu has been used as a foundation towards such a distinct African ethic for the digital era. According to Abubakre, Faik and Mkansi [38], maintaining the values of humbleness, mutuality and goodwill may rein in the competitiveness that characterises modern businesses. They argue for an emerging and adaptive form of Ubuntu, which they call digital Ubuntu, revising traditional values to make a valid and relevant contribution to an information-driven economy. Digital Ubuntu is the integration of communalist values with digital entrepreneurship.

Digital Ubuntu can be embodied in the IS field using Taylor's Ubuntu business ethics principle which has already been referred to above [33]. By replacing "action" with "information system" and extending "parties" to "parties, communities or society at large" the business ethic becomes applicable to IS scenarios. Hence, the fundamental Ubuntu IS ethics principle can be formulated as follows: *An information system is right insofar as it promotes cohesion and reciprocal value amongst people. An information system is wrong insofar as it damages relationships and devalues any individual or group.* The heuristics used to evaluate an information system are:

- Does the information system promote cohesion amongst the parties, communities or society at large?
- Does the information system promote or acknowledge reciprocal value between the parties, communities or society at large?
- Does the information system damage relationships with the various parties, communities or society at large?
- Does the information system devalue any of the parties, communities or society at large?

Support for this ethics core can be found in recent literature on African values and technology. Both Sande [27] (quoted above) and Mujinga [39] have used similar principles in their discussions of the ethical implications of ICT on religion. Mujinga [39] emphasises communality in the principle: “Technology is beneficial to religion because it enhances the communal aspects of religion, and detrimental to religion when it degrades these communal aspects.” The same value underlies Lajul’s [8] formulation of an African biotechnical paradigm as being “centred on the capability building and the social, socioeconomic and environmental contexts”. According to Moyo [20], the use of sex robots promotes individualism and threatens African communitarianism. It is therefore not acceptable in traditional African societies with its high premium on the values of harmony and social cohesion. Regarding the exploitation of data surveillance, Coleman [17] opines that data protection laws are not sufficient to protect users and calls for more reflection to find ways to protect customers’ information. Educating software designers and programmers about the Ubuntu value of respect for other members of the community could go a long way to mitigate the use of such abusive algorithms.

Next, some of the implications of Ubuntu IS ethics are discussed. The field of intellectual property (IP) is an area where Ubuntu IS ethics could make an important contribution. Since the concepts of authorship and IP did not exist in traditional African communities due to their collectivist ethos [40], the Ubuntu ethic should be built out to cover IP issues in the IS field. The following questions should be addressed: What are the implications of Ubuntu as a communalist ethics for the ownership of software applications? How can communities’ intellectual rights be protected from being usurped for economic exploitation by governments or people outside the community?

The use of Ubuntu IS ethics will also have an impact on interaction design and human-computer interaction (HCI). Interface designers need some understanding of the cognitive aspects of interaction to design systems that users will find useful, as well as intuitive to learn and to remember how to use, while also having a pleasurable experience doing so. Cognition theories can be used to inform designers to explain how users interact with applications and to predict how successful they will be in reaching their goals using these apps [41]. Although all humans share the same basic psychological processes, these are often expressed differently in diverse cultures [42]. Assuming that Eurocentric models are universal and simply applying these in African HCI scenarios may therefore alienate indigenous software users. Local philosophical and folk psychology should be used to inform and enrich the conceptual and cognitive frameworks used in localised HCI. This calls for inter- and transdisciplinary collaboration between IS and psychology researchers and practitioners.

This section demonstrated the feasibility of Ubuntu IS ethics. While it should indeed receive its rightful place in IS ethics, there should always be a balancing act to prevent too much emphasis either on individualistic or communalist values. To find this balance, individual and communal rights should be considered and weighed up against each other [43]. Moreover, an equilibrium must be found between the needs of the family, relatives, ethnic groups and smaller cultural communities versus society at large to prevent favouritism and a lack of individual initiatives being regarded as ethically sound in SSAn communities [44].

5 Conclusion

Given the fact that very little research has been done regarding African ethics in the IS field, this essay borrowed ideas from global and African ethics, information ethics and business ethics to serve as pointers for the creation of a unique African IS ethics, called Ubuntu IS ethics. The main contribution of the paper is the formulation of an Ubuntu IS ethics root. Formulating a central ethical value may be regarded as reductionist, but one should take into account that it is just an embryonic phase which could serve as the foundation of a more holistic, detailed and systemised ethic. The core value and its related heuristics may already guide the design and use of IS in communalist environments, albeit to a limited extent.

It is not only important to gain insight into how Ubuntu can enrich IS ethics, but also to study how IS is appropriated to protect African cultures and values such as caring for each other and ensuring a dignified living for all community members [26, 39]. In this regard, Ubuntu IS ethics can also inform global ethics regarding "the role of technology in supporting the resilience of communities" [45]. Borrowing ethical concepts from Ubuntu ethics to formulate a useful Ubuntu IS ethics core demonstrates how African knowledge and epistemology inform and enrich IS ethics, thus answering the essay's research question.

Acknowledgements

The research was supported in part by the National Research Foundation (NRF) of South Africa (grant number 132180) and the Research Professor Support Programme of the University of South Africa (Unisa). The grant holder acknowledges that opinions, findings and conclusions, or recommendations expressed in the article are those of the authors and that neither the NRF nor Unisa accepts any liability whatsoever in this regard. The valuable feedback received from the peer reviewers and editors is also acknowledged with gratitude.

References

1. Bock, A., España, S., Gulden, J., Katharina, J., Nweke, L.O., Richter, A.: The ethics of Information Systems: The present state of the discussion and avenues for future work. In: Proceedings of the 29th European Conference on Information Systems (ECIS 2021), Research-in-Progress Papers, 51, pp. 1–11 (2021). http://hdl.handle.net/10413/14232
2. Kern, C.J., Noeltner, M., Kroenung, J.: The case of digital ethics in IS research – a literature review. In: ICIS 2022 Proceedings, pp. 1–17 (2022). https://aisel.aisnet.org/icis2022/soc_impact_is/soc_impact_is/10
3. Wambsganns, T., Höch, A., Zierau, N., Söllne, M.: Ethical design of conversational agents: Toward principles for a value-sensitive design. In: Wirtschaftsinformatik 2021 Proceedings, no. 2, pp. 1-17 (2021). https://aisel.aisnet.org/icis2022/soc_impact_is/soc_impact_is/10
4. Samuel, M.: Learning and teaching literature: A curriculum development perspective. Alternation: Interdisciplinary Journal for the Study of the Arts and Humanities in Southern Africa 2, 94–107 (1995). https://soe.ukzn.ac.za/?mdocs-file=5595

5. Sartorius, R.: The notion of "development" in Ubuntu. Religion & Development 1, 95–115 (2022). https://doi.org/10.30965/27507955-20220006
6. Otaluka, W.O.: The cultural roots of corruption: An ethical investigation with particular reference to nepotism (PhD thesis). University of KwaZulu-Natal, Pietermaritzburg (2017). http://hdl.handle.net/10413/14232
7. Metz, T.: Toward an African moral theory. Journal of Political Philosophy 15, 321–341 (2007). https://doi.org/10.1111/j.1467-9760.2007.00280.x
8. Lajul, W.: Bioethics and technology: An African ethical perspective. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 189–216. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_12
9. Capurro, R.: Information ethics in the African context. In: Ocholla, D.N., Britz, J.J., Capurro, R., and Bester, C. (eds.) Information Ethics in Africa: Cross-cutting Themes, pp. 7–20. African Centre of Excellence for Information Ethics, University of Pretoria, Pretoria (2013). https://www.researchgate.net/publication/342707988_Information_Ethics_in_Africa_Cross-cutting_Themes
10. Durani, K., Eckhardt, A., Kollmer, T.: Towards ethical design science research. In: ICIS 2021 Proceedings (short paper), no. 3, pp. 1-9 (2021). https://aisel.aisnet.org/icis2021/adv_in_theories/adv_in_theories/3/
11. Pauleen, D.J., Evaristo, R., Davison, R.M., Ang, S., Alanis, M., Klein, S.: Cultural bias in information systems research and practice: Are you coming from the same place I am? Communications of the Association for Information Systems 17, 354–372 (2006). https://doi.org/10.17705/1CAIS.01717
12. Myers, M.D., Chughtai, H., Davidson, E., Tsibolane, P., Young, A.: Studying the other or becoming the other: Engaging with Indigenous peoples in IS research (report on panel discussion on the ethics and politics of engagement with Indigenous peoples in information systems (IS) research at the 40th ICIS, 2019, Munich). Communications of the Association for Information Systems 47, 382–396 (2020). https://doi.org/10.17705/1CAIS.04718
13. Prabhakaran, V., Mitchell, M., Gebru, T., Gabriel, I.: A human rights-based approach to responsible AI. Poster presented at the 2022 ACM Conference on Equity and Access in Algorithms, Mechanisms, and Optimization (EAAMO '22), pp. 1–17 (2022). https://doi.org/10.48550/arXiv.2210.02667
14. Meredith, R., Arnott, D.: On ethics and decision support systems development. In: Proceedings of the 7th Pacific Asia Conference on Information Systems, no. 106, pp. 1562–1575 (2003). http://aisel.aisnet.org/pacis2003/106
15. Gentile, M.C.: Giving voice to values: How to speak your mind when you know what's right. Yale University Press, New Haven, CT (2010).
16. Geeling, S.L., Brown, I.: Conceptualising ethics in the development of social information systems: A grounded theory of social media development. In: ECIS 2020 (research-in-progress papers, no. 5), pp. 1–12 (2020). https://aisel.aisnet.org/ecis2020_rip/5
17. Coleman, D.: Digital colonialism: The 21st century scramble for Africa through the extraction and control of user data and the limitations of data protection laws. Michigan Journal of Race & Law 24, 417–439 (2019). https://doi.org/10.36643/mjrl.24.2.digital
18. Nkohla-Ramunenyiwa, T.: The importance of a neo-African communitarianism in virtual space: An ethical inquiry for the African teenager. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 139–153. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_9
19. Okyere-Manu, B.D.: Shifting intimate sexual relations from humans to machines: An African indigenous ethical perspective. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and

Technology: Questions, Issues, and Approaches, pp. 105–121. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_7

20. Moyo, H.: The death of Isintu in contemporary technological era: The ethics of sex robots among the Ndebele of Matabo. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 123–135. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_8
21. Awuah-Nyamekye, S., Oppong, J.: The use of sex selection reproductive technology in traditional African societies: An ethical evaluation and a case for its adaptation. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 217–228. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_13
22. Morgan, S.N.: Assisted reproductive technologies and indigenous Akan ethics: A critical analysis. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 229–244. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_14
23. Chemhuru, M.: The fourth industrial revolution (4IR) and Africa's future: Reflections from African ethics. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 17–33. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_2
24. Mutula, S.M.: Ethical dimension of the information society: Implications for Africa. In: Ocholla, D.N., Britz, J.J., Capurro, R., and Bester, C. (eds.) Information Ethics in Africa: Cross-cutting Themes, pp. 29–42. African Centre of Excellence for Information Ethics, University of Pretoria, Pretoria (2013). https://www.researchgate.net/publication/342707988_Information_Ethics_in_Africa_Cross-cutting_Themes
25. West, A.: Ubuntu and business ethics: Problems, perspectives and prospects. Journal of Business Ethics 121, 47–61 (2014). https://doi.org/10.1007/s10551-013-1669-3
26. Chirongoma, S., Mutsvedu, L.: The ambivalent role of technology on human relationships: An Afrocentric exploration. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 155–172. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_10
27. Sande, N.: The impact of technologies on African religions: A theological perspective. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 247–261. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_15
28. Naudé, P.: Decolonising knowledge: Can Ubuntu ethics save us from coloniality? Journal of Business Ethics 159, 23–37 (2019). https://doi.org/10.1007/s10551-017-3763-4
29. Konyana, E.G.: Interrogating social media group communication's integrity: An African, utilitarian perspective. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 173–186. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_11
30. Matolino, B.: Values and technological development in an African context. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 73–87. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_5
31. Britz, J.J.: To understand or not to understand: A critical reflection on information and knowledge poverty. In: Ocholla, D.N., Britz, J.J., Capurro, R., and Bester, C. (eds.) Information Ethics in Africa: Cross-cutting Themes, pp. 71–80. African Centre of Excellence for Information Ethics, University of Pretoria, Pretoria (2013).

https://www.researchgate.net/publication/342707988_Information_Ethics_in_Africa_Cross-cutting_Themes

32. Ocholla, D.N.: What is African information ethics? In: Ocholla, D.N., Britz, J.J., Capurro, R., and Bester, C. (eds.) Information Ethics in Africa: Cross-cutting Themes, pp. 21–28. African Centre of Excellence for Information Ethics, University of Pretoria, Pretoria (2013). https://www.researchgate.net/publication/342707988_Information_Ethics_in_Africa_Cross-cutting_Themes
33. Taylor, D.F.P.: Defining Ubuntu for business ethics – a deontological approach. South African Journal of Philosophy 33, 331–345 (2014). https://doi.org/10.1080/02580136.2014.948328
34. Woermann, M., Engelbrecht, S.: The Ubuntu challenge to business: From stakeholders to relationholders. Journal of Business Ethics 157, 27–44 (2019). https://doi.org/10.1007/s10551-017-3680-6
35. Lamola, M.J.: Introduction: The crisis of African studies and philosophy in the epoch of the fourth industrial revolution. Filosofia Theoretica: Journal of African Philosophy, Culture and Religions 10, 1–10 (2021). https://doi.org/10.4314/ft.v10i3.1
36. Lamola, M.J.: Africa in the fourth industrial revolution: A status quaestionis, from the cultural to the phenomenological. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 35–52. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_3
37. Wong-Villacres, M., Alvarado Garcia, A., Maestre, J.F., Reynolds-Cuéllar, P., Candello, H., Iriarte, M., Disalvo, C.: Decolonizing learning spaces for sociotechnical research and design. In: CSCW ’20 Companion, pp. 520–526. ACM (virtual event), New York, NY, October 17–21, 2020 (2020). https://doi.org/10.1145/3406865.3418592
38. Abubakre, M., Faik, I., Mkansi, M.: Digital entrepreneurship and indigenous value systems: An Ubuntu perspective. Information Systems Journal 1–25 (2021). https://doi.org/10.1111/isj.12343
39. Mujinga, M.: Technologization of religion: The unstoppable revolution in the Zimbabwean mainline churches. In: Okyere-Manu, B.D. (ed.) African Values, Ethics, and Technology: Questions, Issues, and Approaches, pp. 263–280. Springer International Publishing, Cham (2021). https://doi.org/10.1007/978-3-030-70550-3_16
40. Kawooya, D.: Ethical implications of intellectual property in Africa. In: Ocholla, D.N., Britz, J.J., Capurro, R., and Bester, C. (eds.) Information Ethics in Africa: Cross-cutting Themes, pp. 43–57. African Centre of Excellence for Information Ethics, University of Pretoria, Pretoria (2013). https://www.researchgate.net/publication/342707988_Information_Ethics_in_Africa_Cross-cutting_Themes
41. Sharp, H., Rogers, Y., Preece, J.: Interaction design: Beyond human-computer interaction. Wiley, Indianapolis, IN (2019).
42. Oppong, S.: The journey towards Africanising psychology in Ghana. Psychological Thought 9, 1–14 (2016). https://doi.org/10.5964/psyct.v9i1.128
43. Geeling, S.L., Brown, I.: Towards ethical social media practice: A grounded theory for analyzing social media platform ethics. SSRN Electronic Journal 1–17 (2021). https://doi.org/10.2139/ssrn.3942451
44. Lutz, D.W.: African Ubuntu philosophy and global management. Journal of Business Ethics 84, 313–328 (2009). https://doi.org/10.1007/s10551-009-0204-z
45. Leal, D.D.C., Krüger, M., Teles, V.T.E., Teles, C.A.T.E., Cardoso, D.M., Randall, D., Wulf, V.: Digital technology at the edge of capitalism. ACM Transactions on Computer-Human Interaction 28, 1–39 (2021). https://doi.org/10.1145/3448072

Artificial Intelligence Impact on the realism and prevalence of deepfakes

Oyena Mahlasela [0009-0006-4188-6145] Errol Baloyi [0009-0009-6959-3485]

Nokuthaba Siphambili[0009-0006-6091-9838] and Zubeida C. Khan[0000-0002-1081-9322]

Council of Scientific and Industrial Research, Pretoria, South Africa
omahlasela@csir.co.za, ebaloyi2@csir.co.za,
nsiphambili@csir.co.za, zdawood@csir.co.za

Abstract. Deepfakes, synthetic media manipulated by Artificial Intelligence (AI), have become a growing concern in the information landscape. This paper explored the impact of AI on the realism and prevalence of deepfakes. Therefore, this study examined how AI advancements in machine learning and generative models have facilitated the creation of increasingly convincing deepfakes. The analysis looked at the rise of hyper-realistic deception and the societal impact of deepfakes. In addition to recognizing the challenges, a framework was developed for the detection of deepfakes. Finally, this study discussed the potential mitigation strategies, such as the development of deepfake detection tools and fostering media literacy.

Keywords: Deepfakes, Artificial Intelligence, Machine Learning, Generative Models, Disinformation, Media Literacy

1 Introduction

Media manipulation has been around for decades. However, the emergence of deepfakes presented a significant threat due to their hyper-realistic synthetic media generated by Artificial Intelligence (AI) techniques such as machine learning (ML) algorithms and generative models [1]. The rise of deepfakes stemmed in 2017 when a controversial Reddit user shared AI-generated adult images of celebrities; this incident was possibly created by stitching together shallow ML models for facial replacement and existing celebrity footage [2] [3]. The term "deepfakes" originated from an anonymous Reddit user, known as 'deepfakes', who created and shared the first deepfake; this sparked widespread interest within the Reddit community and led to a proliferation of fake content [42]. Subsequently, individuals began to utilise deepfake technology for various purposes, including creating humorous videos for sharing among friends, producing comedy shows, crafting political commentary, and even developing immersive artworks and cultural experiences [43]. However, the malicious use of deepfakes escalated over time, with notable targets including celebrities, actors, singers, and politicians. Even though the early deepfakes were less sophisticated, such as the 2017 deepfake incident. However, it introduced the concept of manipulating reality with AI. Igniting discussions around technology misuse and ethical concerns. This encouraged conversations about the need for safeguarding to prevent the application of deepfakes for malicious purposes, like creating fake news [4][5][6].

However, between 2018 and 2019, deepfake technology became easily accessible due to the availability of pre-trained models and open-source tools [3][7]. This resulted in people having a platform to create deepfakes because of the availability of technological capabilities. There was also a rise in malicious use, targeting mostly prominent people and creating social unrest [8]. To date, accessing deepfake technology is painless, and credit should be given to user-friendly interfaces and cloud-based processing, which democratise deepfake creation [4][8]. Lowing the barriers to entry by reducing the technical expertise needed to understand AI, ML, and potential programming skills. Thus, creating deepfake media only requires uploading targeted images and source footage, and the platform takes care of the complex computations behind the scenes [3]. Cloud platforms also enable tasks to be performed with increased speed and efficiency since they offer pre-built templates and tools that exist for specific tasks, such as cloning the voice or facial swaps [8].

Therefore, the advancement of cloud-based deepfakes creation tools and audio deepfakes can superimpose a person's voice and likeness, blurring the lines between fabrication and reality [8]. There is also this new relationship between deepfakes and AI mainstreaming, as deepfakes technology uses deep learning algorithms as the foundation for complex video and audio data patterns to generate realistic fakes [1][7]. For example, Generative adversarial networks (GANs) are a type of deep learning algorithm that has been mostly effective in creating deepfakes [1]. GANs work by opposing two neural networks against each other: one network generates fake data, while the other network tries to distinguish fake data from real data [1]. As a result, this could create realistic videos of people saying things they have never communicated. Thus, it is imperative to consider a proactive stance in developing deepfake detection methods using AI to identify manipulation signatures. This study investigates the realism behind deepfakes, the threats, and the disinformation and erosion of trust. Then, a framework was constructed for the mitigation of deepfake threats.

The remainder of the paper is structured as follows. The methodology for the study is presented in Section 2, followed by a literature review in Section 3. A discussion is provided in Section 4. In Section 5, a solution framework is presented. Section 6 concludes the paper with directions for future work.

2 Methodology

This research study applied a systematic literature review (SLR) to examine the impact of deepfakes on the AI era [14]. Research publications were synthesised to understand the realism, threats, and disinformation of deepfakes. Thirty-five publications written between 2022 and 2024 were included in the final review. This was to investigate the rise of deepfakes in the past three years as generative AI emerged into mainstream technology. Scholarly databases that were used to retrieve the search were Science Direct, Springer Link, Scopus and Google Scholar search engine. The keywords applied in various databases were (“deepfakes”, “misinformation”, “disinformation”, “fake news”, OR “false information”, AND “generative artificial intelligence) for more comprehensive results. The selection criteria excluded publications that were not written in English, any duplicates of studies were also excluded, and finally, publications that were not accessible were also excluded. The final publications were included in the PRISMA diagram as illustrated in Fig.2 below:

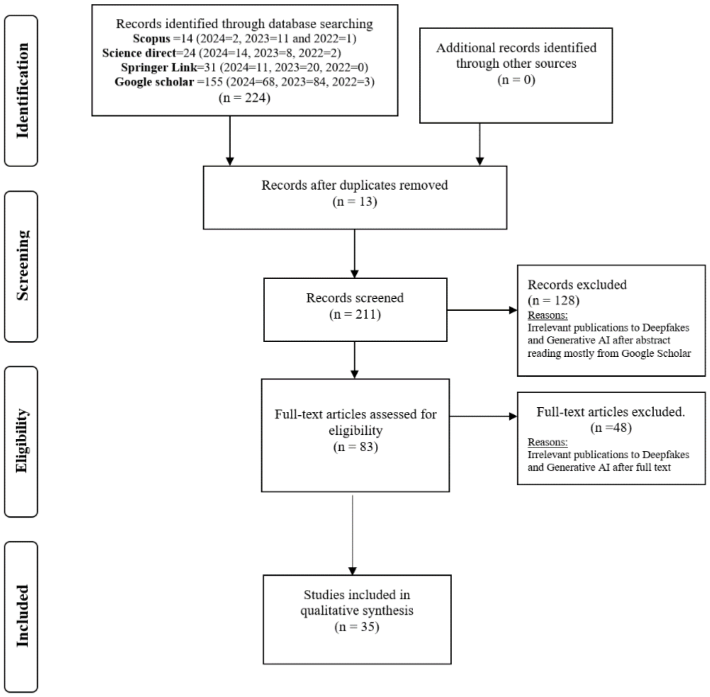

Identification
Screening
Eligibility
Included
Records identified through database searching
Scopus =14 (2024=2, 2023=11 and 2022=1)
Science direct=24 (2024=14, 2023=8, 2022=2)
Springer Link=31 (2024=11, 2023=20, 2022=0)
Google scholar =155 (2024=68, 2023=84, 2022=3)
(n = 224)
Additional records identified through other sources
(n = 0)
Records after duplicates removed
(n = 13)
Records screened
(n = 211)
Records excluded
(n = 128)
Reasons:
Irrelevant publications to Deepfakes and Generative AI after abstract reading mostly from Google Scholar
Full-text articles assessed for eligibility
(n = 83)
Full-text articles excluded.
(n =48)
Reasons:
Irrelevant publications to Deepfakes and Generative AI after full text
Studies included in qualitative synthesis
(n = 35)

Fig1: PRISMA Diagram

The research question addressed in this study was:

- RQ: What is AI's impact on advancing deepfakes?

3 Literature review: AI: The engine behind deepfake realism

Deepfakes have evolved in recent years. To date, deepfakes leverage the power of AI techniques, like ML algorithms, to analyse immense amounts of data, such as audio recordings and videos, by learning characteristics of the person's speech, patterns, and appearance [13]. Regarding generative models, deepfakes use learned information to analyse and synthesise realistic, new media for the targeted person [14]. Therefore, continuous advancement of generative AI deepfakes enhances their quality, making it increasingly difficult to distinguish fakes from authentic content. Thus, this study evaluated some of the deep generative models that can contribute to the advancement of more realistic deepfakes.

3.1 Variational Autoencoders (VAEs)

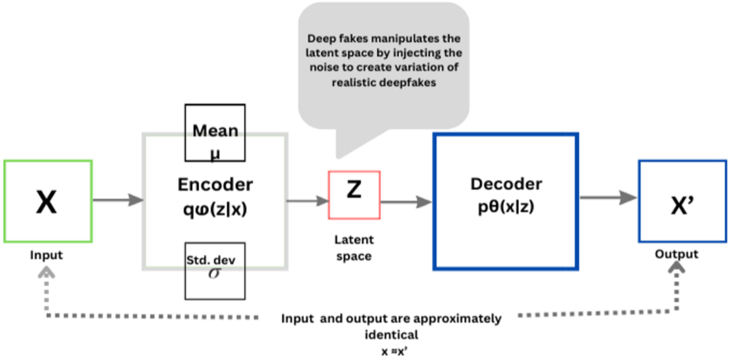

Deep fakes manipulates the latent space by injecting the noise to create variation of realistic deepfakes
Mean
μ
X
Input
Encoder
qω(z|x)
Std. dev
σ
Z
Latent space
Decoder
pθ(x|z)
X'
Output
Input and output are approximately identical
x ≈x'

Fig.2. Variational Autoencoders Model [48]

The variational autoencoders are generative models that use unsupervised learning; they compress and restrict data into a latent space [5]. The VAE models in Fig.***22***. illustrate that the input (x) is entered, and it is encoded in the probabilistic zone $q_\varphi(z|x)$; the data then is compressed in the latent space (z) and thereafter decoded p(x|z). The output is then represented by x'. VAEs generate variations of data based on the latent space (z). For example, VAEs can generate new portrait images with different expressions. Thus, the misuse of VAEs can lead to the manipulation of the deep generative models' training data, which can potentially generate biased outputs towards fakes. Deepfakes manipulation in VAEs involves the modification of latent space represented by (z).

The encoder $q_\varphi(z|x)$ can be used to create the latent represented by (z_target) for the targeted video, which is usually the person targeted for manipulation, and the (z_source) is the source video, the person whose action is copied. The latent space manipulation vectors are usually manipulated through linear interpolation, which can be done by creating a blend between the two latent vectors to generate an intermediate representation for creating transitions [19]. The other way latent space manipulation can occur is by injecting noise into the latent space to create variations, improving the deepfakes' realism [18]. Both the manipulation of (z_target and z_source) creates a new latent space to form (z), which becomes key when encoding the desired action and appearance for deepfakes. However, the decoder pθ(x|z) is the one used to decode the manipulated latent space (z) to generate deepfake media. Nonetheless, manipulating the VAEs can lead to unrealistic outputs and unintended artefacts. It is the type of generative model that is not widely used to create deepfakes but is often used in conjunction with GANs. Furthermore, it is important to note that VAEs do not always produce high-quality deepfakes images. However, the commonly used generative model is GANs.

3.2 Generative Adversarial Networks

GANs are also a type of deep learning that can be used to generate new data [20]. GANs consist of two neural networks: the generator that generates new data and the discriminator, which classifies or distinguishes fakes/ real data, as shown in Fig. ***3***3 [49]. The two neural networks are then trained iteratively in an adversarial process to the point where the generator improves and begins to create realistic fakes while the discriminator improves at spotting the fakes [21][22]. This competitive process drives both neural networks to improve performance. Ideally, this extensive training should encourage the generator to become skilled at producing new data that can fool the discriminator, meaning the generated data should statistically be similar to the real data [22]. The application of GANs includes creating realistic images, restoring damaged paintings, and generating new data.

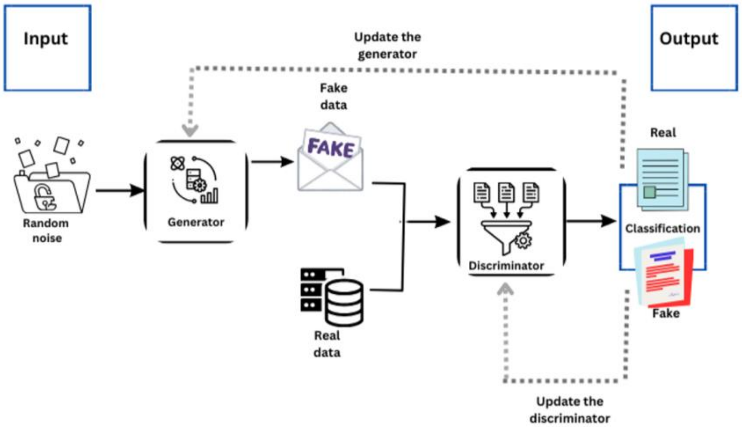


Fig. 3. Generative Adversarial Networks [49].

Thus, GANs, when misused, can create realistic fakes because they excel at learning patterns of data, such as faces in videos. Furthermore, continuous improvements between the generator and the discriminator help create increasingly difficult-to-detect fakes [23]. The other thing that can encourage the use of GANs in creating deepfakes is their adaptability; GANs can be trained in various data types [24]. This means GANs can generate deepfakes of images, videos, or audio since they can also manipulate speech patterns [24].

3.3 Autoregressive Models (ARs)

Autoregressive Models (ARs) are statistical models used to forecast time series data. They are particularly used when future values of series can be used to predict past values [25]. For example, predicting the next word in a sentence. The AR model consists of time series data, which are data points collected at regular intervals over time; the lag refers to the time difference between data points [26]. Then, there is aggression, which predicts future values by considering the series of past values, and this basic formula formulates the AR model as follows:

$$yt=c+\varpi_1\, y(t-1) + \varpi_2\, y(t-2) +\cdots+ \varpi pyt-p+ \varepsilon t,$$

The (εt) is the error term representing the data's unpredictable random noise. The coefficients are indicated by phi (ϖ) and are assigned to the past values y(t-1) to y(t-2). The constant term is (c), and (yt) represents the value to be predicted at time (t). ARs can be used to create realistic speech patterns in platforms such as chatbots. Therefore, if misused, ARs can provide manipulated training data to the AI generative models. Potentially biasing the generated outputs towards fakes [27]. Hence, ARs can contribute to deepfakes through trained speech synthesis to generate fake audio or text.

3.4 Flow-based Models (FBMS)

The flow-based models (FBMs) are another generative model that transforms a simple noise that is distributed into complex data. FBMs, as illustrated in Fig.4. can generate new data by revising the flow, and this can be done by learning the patterns of images well enough to generate realistic, new images [28]. Thus, FBMs rely on normalising flow, which are flows with sequences of invertible mathematical functions $(f^{-1}(z))$. The $(f^{-1}(z))$ function can transform simple probability distribution, e.g. random noise (x), into a more complex distribution of the data image (x'). Furthermore, each step in the flow is invertible, meaning FBMs can generate old data and assist the users in understanding how the model arrived at that data by reversing the flow. Therefore, each function modifies the data to bring it closer to the targeted distribution (x'). Thus, composing multiple flows allows the model to capture more complex data patterns and make them perform well in functions like generating realistic images, predicting sequences of data points, and identifying data points that deviate from the norm (e.g. anomaly detection). Fig 3 shows the flow-based model.

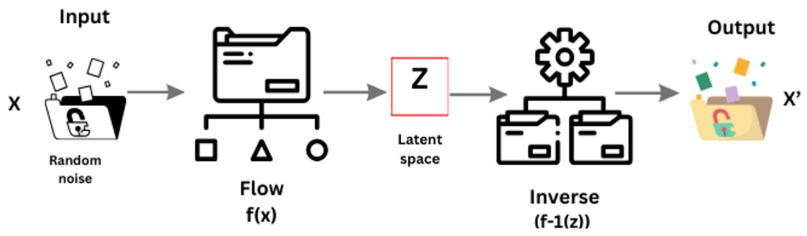


Fig.4. Flow-based Model [48]

However, even though FBMs can generate realistic images. For example, they can generate realistic 3D models of objects from 2D images [30]. They are not widely used to create deepfakes like GANs, but they can assist in the generation of deepfakes. FBMs have the invertible function ($f^{-1}(z)$), which enables them to reverse their transformation; this could be useful for creating more refined expressions on the targeted face in deepfakes. Moreover, FBMs are known to be efficient data generators [29]. They are not as dominant as GANs in creating deepfakes, and training FBMs for deepfakes might be a challenge. The other interesting thing about FBMs is that their invertible function can be used to detect deep fakes. However, more research is needed to explore this.

3.5 Diffusion Models (DMs)

Diffusion models (DMs) are also a type of generative model that can be used to create new images. DMs are known for gradually adding noise to the image in small steps and turning it into a random noise as shown in Fig. ***5***5. With DMs, the models are trained to reverse the process from a noisy version back to the original clear image [31][32]. Therefore, the model can create entirely new images from noise and reverse the noise process. It is, however, important to learn the patterns and steps needed to create new and realistic data from scratch. However, DMs are known for producing very realistic detailed data, images in particular. Their interpretable latent space enables them to navigate the noise to reach the desired data point, which can also be helpful in providing insight into the data itself. Furthermore, when compared to other generative models, DMs are less likely to memorise training data. Thus, making them less prone to overfitting [33][34].

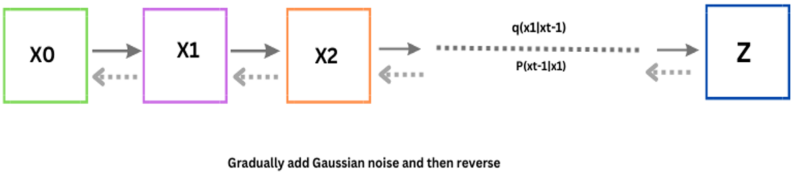


Fig. 5. Diffusion model [48]

DMs can be applied in various domains, such as text-to-image synthesis, image generation and editing, and even in more complex domains like scientific data augmentation and music composition. However, since DMs excel at creating realistic images, they can be ideal for generating deepfakes that are difficult to distinguish from real videos or images [36][35][34]. Furthermore, DMs enable text prompts to guide the generation process, which can encourage deepfakes creators to use text descriptions to manipulate features such as facial expressions. Thus, after understanding the engine behind creating deepfakes, it was important to investigate the threats alongside deepfake creation.

The literature review revealed that most studies focus on the capabilities of generative models like GANs in creating realistic deepfakes [20]-[22]. Human intervention and media literacy in mitigating the impact of deepfakes is not considered in studies. There is a need for strategies that empower individuals to critically assess the authenticity of media.

4 The Prevalence of Deepfakes: A Looming Threat

Deepfakes are created to mislead people with wrong information using fake videos and images. AI has increased the use of deepfakes as more and more tools are being used to create deepfakes. The ease of use of deepfakes because of AI has created havoc, as anyone with access to generative models can generate deepfakes by manipulating images and videos [10]. Furthermore, AI has made it easier for threat actors with limited expertise to generate deepfakes. The increasing use of deepfakes is seen in cases such as elections. Deepfakes are a threat to elections and democracy in most countries as threat actors use fake images and videos of politicians to spread their agenda and misinform the public about a certain politician [10][11][13]. Spreading propaganda of hate speech or autocratic views. People who may not be able to identify deepfakes are more susceptible to false beliefs; they are prevented from acquiring the right information and knowledge due to reliance on the videos or images they may be exposed to [11].

As deepfakes become more sophisticated, people may lose trust in videos, images, text, audio or any other media, as it may become difficult to distinguish between fakes and real data. Threat actors may also use deepfakes to gain access to individuals' credentials for financial gain by accessing sensitive data and systems. When a criminal has access to one's credentials, they can manoeuvre around networks and systems to perform criminal activities and cause operational disruption. Furthermore, organisations may be at risk of financial losses if employees are not trained to identify deepfakes, which may result in exposure to security vulnerabilities. Deepfakes may also cause reputational damage to the dispensation of news, and this can be a threat to journalism as it may impede confusion in distinguishing between fake and real news [13].

The other looming threat is the use of deepfakes for unethical use. Deepfakes are used for various purposes, such as the creation of pornographic content of innocent people without their permission, and this may result in reputational damage and instil fear in the victim [15]. In as much as deep fakes can be used harmlessly, like creating a fake picture on social media pretending to be on a trip to Paris while you are at home, deepfakes can also be used maliciously for ransomware, disinformation and sometimes to threaten people's privacy and security.

Open-source AI tools, along with emerging technologies, are facilitating innovative approaches to deepfake deception. The engine behind deepfakes realism is driven by deep learning technologies. Most studies pointed to GANs as the driving generative model behind deep fakes [22][49]. The VAEs were the other model that was found to be coming along in deepfake creation, but they were mostly used in collaboration with GANs since they are expensive to train [18]. Other generative deep models, ARs and FBMs, were viewed as supporting models that could assist in perfecting deepfake science [26][29]. However, the diffusion model was another generative model that could create even more realistic fakes since it excels at creating realistic images that are difficult to distinguish from real images and videos [48].

As emphasised in the study, these technologies can generate seemingly authentic videos of individuals without their consent [40]. With advancements in AI and ML, deepfake techniques provide automated methods for producing synthetic content that is increasingly difficult for humans to distinguish, thereby expanding the potential for deception across various media formats, including images, videos, and audio recordings [42]. The concept of deepfakes became widely accessible to users in 2017, although the movie industry has been using similar techniques for some time. For example, movies have used computer-generated imagery (CGI) to portray deceased actors. However, creating such deepfakes traditionally required significant resources, expertise, and specialised software. Today, many of these capabilities are available on smartphones, and many of the tools are free [42][45].

The use of deepfakes for pornography is a significant concern at a personal level. Victims and pornographic performers often do not consent to the use of their images in such a manner, and even though videos can be removed, they often continue to circulate widely in less reputable corners of the internet. Studies have shown that a vast majority (96%) of online deepfakes are pornographic [41]. Another critical issue is the rise of synthetic child pornography. Deepfakes are especially troubling in the realm of child sexual abuse because they can be used to create new online child sexual abuse material from existing material. This means that creators could potentially produce images of children being abused or fabricate material using images of children who have not been victims of actual abuse. Deepfake technology allows a creator to superimpose a person's face onto another's body in a video, enabling the use of images from various online sources, such as social media, to be manipulated in this manner [43].

Today, distinguishing between authentic and artificially created videos is increasingly challenging as deepfake technology improves rapidly, making the generated content more believable [42]. An illustrative example of this advancement is evident in a fabricated picture that depicts two former heads of state, Barack Obama and Angela Merkel, seemingly enjoying a leisurely stroll at the beach. Despite their convincing appearance, these images were created by a visual artist named Julian and shared on Instagram. Similarly, other deepfake images, including those depicting former President Trump being arrested, have also gained viral attention. These manipulated images have the potential to evoke strong reactions, particularly among individuals who are heavily engaged and consume social media news. Furthermore, conspiracy theorists may also exploit both sets of images to advance their agendas [45][46].

At a government level, such manipulated images can have serious implications and can be used to influence political outcomes [41][40]. Presently, already there is a witness of the usage of deepfakes in political communication; for instance, during a state-level election campaign in India in February 2020, a political party admitted to spreading manipulated campaign deepfakes via WhatsApp [44]. Furthermore, the Flemish Socialist Party released a deepfake video depicting Donald Trump seemingly advocating for Belgium to withdraw from the Paris Climate Accords. Despite obvious signs of manipulation, such as the speaker's mouth being out of sync with the facial expressions, some individuals failed to recognise it as fake [41]. One early influential deepfake, "Synthesizing Obama," demonstrated the power of AI and deep learning by convincingly lip-syncing audio to existing footage of Obama [42]. The deepfake technology raises the possibility of watching a leader convincingly deliver a speech attributed to another leader, or vice versa, blurring the lines between reality and fabrication.

The use of AI to generate images of people without their consent is a serious ethical concern. However, the rapid development and prevalence of deepfake technology have left legal systems lagging existing laws on content regulation have not kept pace with technological advancements [43][45]. According to [37], the heavy reliance on social media platforms has exacerbated the situation. [37] study showed that social media has evolved from a communication tool to a powerful weapon capable of shaping public opinion, influencing perceptions, and steering events. This was evident in the recent Russian-Ukraine conflict, where cyberspace played a crucial role. Propaganda and misinformation were spread through social media during the conflict, with videos of explosions going viral. Additionally, Ukrainian President Zelenskyy was the target of a deepfake video, created using AI to make it appear he said something he did not. Furthermore, [37] notes that the strategic use of disinformation as a weapon of war aims to impede international relations, erode public trust in leaders and institutions, and undermine opposing political ideologies.

The rise of AI has drawn the interest of hackers and cyber attackers seeking new ways to exploit and manipulate technology for financial gain [39]. At the organisational level, phishing emails have long been a persistent threat. Phishing aims to deceive victims into making errors or disclosing sensitive information by enticing them to open documents, files, or emails, visit websites, or grant access to systems or services [38]. While social engineering methods are commonly used to gain initial access, they can also be employed in the later stages of an incident or breach. Examples include business email compromise (BEC), fraud, impersonation, counterfeiting, and, more recently, extortion. These phishing schemes have evolved with the introduction of AI and deepfakes, marking a significant shift in the creation of fake content [40].

For instance, a finance worker at a multinational corporation was tricked into transferring $25 million to fraudsters who used deepfake technology to impersonate the company's chief financial officer in a video conference call. Initially suspicious of a phishing email from someone appearing to be the UK-based CFO, the worker's doubts were eased during the video call when other participants sounded like familiar colleagues [47]. This incident is an example of BEC, a sophisticated scam that targets businesses using social engineering [38]. Another similar incident involved the CEO of a UK energy firm who received a call from his boss, the CEO of the company's German parent company, requesting urgent funds transfer to a Hungarian supplier. Recognising the voice and accent, the CEO transferred funds but later became suspicious and avoided a second payment request [12][40].

Malicious synthetic media poses an emerging threat to organisations and financial markets, capable of spreading misinformation, disinformation, and defamation that can severely damage an employee's or an organisation's reputation [40]. This study examined various deepfake techniques, the advancing sophistication of AI algorithms, and their combined impact on the realism and prevalence of fake media. To offer insights into the current state of deepfakes, a framework for detecting and mitigating their harmful effects is discussed in the following section.

5 Framework for detecting deepfakes.

Combating the deepfake threat requires a multi-layered approach. This study examined the impact of AI on the advancement of deepfakes. Therefore, it was clear that mitigating deepfakes is not only limited to technological detection. However, advancements in deepfake detection, using AI techniques to identify manipulated content, are crucial. Additionally, fostering media literacy through educational initiatives is essential to equip individuals with the skills to evaluate online content critically. This study notes that there are limitations to deepfake detection by technological detection only, more so in a South African context where there is a digital divide and media literacy concerns. The dimensions of the proposed framework shown in ***Fig 6*** are presented here.

5.1 Legislative mitigation

Governmental organisations need to draft legislation to impose sanctions on the development and dissemination of deepfakes. It is crucial to notify citizens that such regulation can be drafted to allow for freedom of speech but to prevent misuse and harm of others. In an effort to get trust from citizens on regulations, governmental departments could host panel discussions and gather feedback to ensure the regulations are clear, enforceable, and respect freedom of expression. To some extent, social media platforms should also enforce terms of usage and guidelines on disseminating deepfakes through their platforms by enforcing bans and removal of content.

5.2 Technological mitigation

Technological advancements are the most accurate way to identify deepfakes by using AI techniques, forensics, and hybrid approaches to detect anomalies and patterns within deepfakes. However, there is a challenge that most AI tools are not inclusive to cater to local languages, dialects and accents in the South African sense. Natural Language Processing (NLP) researchers need to collaborate and contribute to technological advancements in this area.

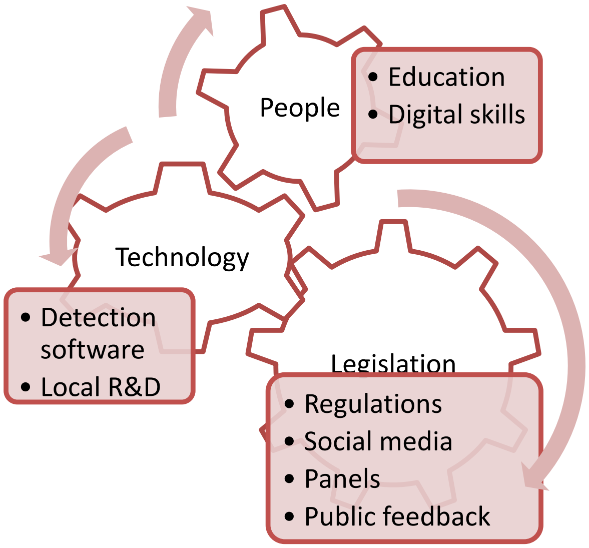


Fig 6: The framework for detecting deepfakes; arrows represent collaboration.

5.3 Human-centric mitigation

Human-centric mitigation is required to empower citizens to assess the credibility of videos. Consider the case of Singapore, where the government seeks to encourage adults to attend technology schools. This is towards an effort to equip its citizens with the digital skills needed to thrive in the modern world. This includes skills relevant to deepfakes, such as media literacy and critical thinking in a digital environment. While Singapore's initiative focuses on broader digital skills development, the core skills learned can be directly applied to the fight against deepfakes in South Africa. Another route is to have a crowdsourced platform where citizens can flag suspicious online content. Vetted volunteers can then analyse the flagged content using online resources and basic verification tools.

5.4 Collaboration

Collaboration is the key driver that binds the legislative, technological, and human-centric aspects of mitigating deepfakes. Clear and enforceable regulations, informed by public discussions and feedback, can provide the framework for social media platforms to remove malicious deepfakes. Meanwhile, researchers can develop AI tools tailored to South African languages and dialects through collaboration with local experts. Furthermore, crowdsourcing platforms, where citizens flag suspicious content, can be most effective when partnered with the technological sector to train volunteers and provide them with the necessary verification tools.

5.5 Use-case

In this scenario, demonstration of the framework is shown when an executive assistant of a company in the mining domain receives an email purportedly from the CEO, instructing them to transfer a sum of money to an offshore account for an emergency. The email contains a deepfake video, emphasising its time-sensitive nature.

Legislative mitigation: Upon discovering the incident, the company immediately reports the deepfake fraud to relevant authorities. Legislators work swiftly to enact laws that explicitly prohibit the creation and dissemination of deepfakes for fraudulent purposes. Public awareness campaigns are launched to educate organisations about the risks posed by deepfake technology and the legal repercussions for its misuse.

Technological mitigation: The company invests in advanced AI-powered detection tools specifically tailored to identify deepfake videos and audios. These tools utilise sophisticated algorithms capable of detecting anomalies and inconsistencies indicative of manipulation. Additionally, collaborations with AI researchers and natural language processing experts are initiated to enhance the tools' effectiveness, particularly in detecting deepfakes in local languages and accents prevalent in South Africa.

Human-Centric Mitigation: Recognising the importance of human intervention in detecting and preventing deepfake fraud, the company launches a comprehensive training program for its employees. The program focuses on enhancing digital literacy and critical thinking towards unsolicited requests, especially those accompanied by multimedia content. Employees are trained to verify the authenticity of messages, particularly those involving financial transactions, through multiple channels and cross-referencing with known communication patterns.

Collaboration: The company collaborates with regulatory bodies, law enforcement agencies, and other institutions to share information and best practices in combating deepfake fraud. Cross-industry partnerships are formed to develop standardised protocols for verifying the authenticity of electronic communications, particularly those involving sensitive financial transactions.

Outcome: Through the concerted efforts outlined in the framework, the company successfully mitigates the threat of deepfake fraud and safeguards its financial transactions against future incidents. By combining legislative, technological, and human-centric approaches in a collaborative manner, the company demonstrates its commitment to combating emerging threats posed by deepfake technology.

6 Conclusion & Future Work

The explosion of deepfakes presents a significant challenge to the information landscape. As a result, this study drew from an array of interdisciplinary perspectives, including AI, ML, ethics, and cybersecurity, to comprehensively paint a picture of the deepfake landscape and the impact of AI on the realism and prevalence of deepfakes. Furthermore, it highlighted how advancements in ML and generative models have facilitated the creation of increasingly convincing deepfakes, raising concerns about disinformation and erosion of trust. Moreover, this study concludes that with the increasing accessibility of technology and the ease of distribution through social media platforms, the creation and distribution of deepfake content will continue to rise. The proposed framework will aid in addressing the mitigation of criminal activities via deepfakes. This research will benefit a wide range of stakeholders, including policymakers, social media organisations, and most importantly, the public. There is still a need for vigilance around deepfakes. Promoting transparency around deepfake creation and educating users on how to critically evaluate online content need to be performed using the framework. In future, the authors propose the creation and distribution of deepfake content to be carried out, together with an analysis of how much impact deepfake content has on users. Therefore, researchers should invest in developing AI detection platforms that will flag suspicious online content, especially to cater to local languages.

References

1. Sharma, M. and Kaur, M.: A review of Deepfake technology: an emerging AI threat. Soft Computing for Security Applications: Proceedings of ICSCS 2021, pp.605-619. (2022)
2. Culver, C.: (Ed.). Griffith Review 79: Counterfeit Culture (Vol. 79). Griffith Review. (2023)
3. Grothaus, M.: Trust No One: Inside the World of Deepfakes. Hachette UK. (2021)
4. Chesney, B., & Citron, D.: (2019). Deep fakes: A looming challenge for privacy, democracy, and national security. Calif. L. Rev., 107, 1753. (2019)
5. Krishna, D.: Deepfakes, online platforms, and a novel proposal for transparency, collaboration, and education. Rich. JL & Tech., 27, 1. (2020).
6. Collins, A.: Forged authenticity: governing deepfake risks. (2019).
7. Helmus, T. C.: Artificial Intelligence, Deepfakes, and Disinformation.
8. Jones, V. A.: Artificial intelligence enabled deepfake technology: The emergence of a new threat (Doctoral dissertation, Utica College). (2020).
9. Fallis, D.: The epistemic threat of deepfakes. Philosophy & Technology, 34(4), 623-643. (2021).
10. Milmo, D.: The Guardian, https://www.theguardian.com/world/2024/feb/05/hong-kong-company-deepfake-video-conference-call-scam, last accessed 2024/18/03).
11. Cinar, B.: Deepfakes in Cyber Warfare: Threats, detection, techniques and counter-measures. Asian Journal of Research in Computer Science, 16 (4), 178–193 (2023).
12. Bansal, U.: A review on Ransomware attack, 2021 2nd International Confer-ence on Secure Cyber Computing and Communications (ICSCCC), May 2021. doi:10.1109/icsccc51823.2021.9478148. (2021)
13. Graber-Mitchell, N.: Artificial illusions: Deepfakes as speech, Intersect, 14(3), Avail-able at SSRN: https://ssrn.com/abstract=3876862 (2021).
14. Shahzad, H. F., Rustam, F., Flores, E. S., Luis Vidal Mazon, J., de la Torre Diez, I., & Ashraf, I.: A review of image processing techniques for deepfakes. Sensors, 22(12), 4556. (2022).
15. Pawelec, M.: Deepfakes and democracy (theory): How synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. Digital society, 1(2), p.19. (2022).

16. George, A. S., & George, A. H.: Deepfakes: The Evolution of Hyper realistic Media Manipulation. Partners Universal Innovative Research Publication, 1(2), 58-74. (2023).
17. Bontcheva, K., Papadopoulous, S., Tsalakanidou, F., Gallotti, R., Krack, N., Teyssou, D., ... & Verdoliva, L.: Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities. European Digital Media Observatory. (2024).
18. Zhao, Y., Liu, B., Ding, M., Liu, B., Zhu, T., & Yu, X.: Proactive deepfake defence via identity watermarking. In Proceedings of the IEEE/CVF winter conference on applica-tions of computer vision (pp. 4602-4611). (2023).
19. Nowroozilarki, Z., Mortazavi, B. J., & Jafari, R.: Variational autoencoders for biomedical signal morphology clustering and noise detection. IEEE Journal of Biomedical and Health Informatics (2023).
20. Porkodi, S.P., Sarada, V., Maik, V. et al.: Generic image application using GANs (Generative Adversarial Networks): A Review. Evolving Systems 14, 903–917 (2023).
21. Yates, M., Hart, G., Houghton, R., Torres, M.T. and Pound, M.: Evaluation of synthetic aerial imagery using unconditional generative adversarial networks. ISPRS Journal of Photogrammetry and Remote Sensing, 190, 231-251 (2022).
22. Di Zio, S., Calleo, Y. and Bolzan, M.: Delphi-based visual scenarios: an innovative use of generative adversarial networks. Futures, 154, 103280 (2023).
23. He, X., Chang, Z., Zhang, L., Xu, H., Chen, H. and Luo, Z.: A survey of defect detection applications based on generative adversarial networks. IEEE Access, 10, 113493-113512 (2022).
24. Wu, A.N., Stouffs, R. and Biljecki, F.: Generative Adversarial Networks in the built environment: A comprehensive review of the application of GANs across data types and scales. Building and Environment, 223, 109477 (2022).
25. Liu, Z. and Yang, X.: Cross validation for uncertain autoregressive model. Communications in Statistics-Simulation and Computation, 51(8), 4715-4726 (2022).
26. Nguyen, T.T., Nguyen, Q.V.H., Nguyen, D.T., Nguyen, D.T., Huynh-The, T., Nahavandi, S., Nguyen, T.T., Pham, Q.V. and Nguyen, C.M.: Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 223, p.103525 (2022).
27. Vikranth, K., Nethravathi, P.S., Krishna Prasad. K.: Price prediction system – a predictive data analytics using arima model. Ictact Journal on soft computing, 14(3), 3304-3310 (2024).
28. Ravanmehr, R. and Mohamadrezaei, R.: Deep Generative Session-Based Recommender System. In Session-Based Recommender Systems Using Deep Learning (pp. 119-169). Cham: Springer Nature Switzerland (2023).
29. Du, H., Niyato, D., Kang, J., Xiong, Z., Zhang, P., Cui, S., Shen, X., Mao, S., Han, Z., Jamalipour, A. and Poor, H.V.: The age of generative AI and AI-generated everything. arXiv preprint arXiv:2311.00947 (2023).
30. Luleci, F., Catbas, F.N: A brief introductory review to deep generative models for civil structural health monitoring. AI Civ. Eng. 2, 9 (2023). https://doi.org/10.1007/s43503-023-00017-z
31. Moser, B.B., Shanbhag, A.S., Raue, F., Frolov, S., Palacio, S. and Dengel, A.: Diffusion Models, Image Super-Resolution And Everything: A Survey. arXiv preprint arXiv:2401.00736 (2024).
32. Wu, Q., Liu, Y., Zhao, H., Kale, A., Bui, T., Yu, T., Lin, Z., Zhang, Y. and Chang, S.: Uncovering the disentanglement capability in text-to-image diffusion models. In Proceed-ings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 1900-1910 (2023).

33. Seider, N.A., Adeyemo, B., Miller, R., Newbold, D.J., Hampton, J.M., Scheidter, K.M., Rutlin, J., Laumann, T.O., Roland, J.L., Montez, D.F. and Van, A.N.: Accuracy and reliability of diffusion imaging models. NeuroImage, 254, p.119138 (2022).
34. Lutati, S. and Wolf, L.: Ocd: Learning to overfit with conditional diffusion models. In International Conference on Machine Learning (pp. 23157-23169). PMLR (2023).
35. Lorenz, P., Durall, R.L. and Keuper, J.: Detecting images generated by deep diffusion models using their local intrinsic dimensionality. In Proceedings of the IEEE/CVF Inter-national Conference on Computer Vision. pp. 448-459 (2023).
36. Bammey, Q.: "Synthbuster: Towards Detection of Diffusion Model Generated Images," in IEEE Open Journal of Signal Processing, vol. 5, pp. 1-9 (2024). doi: 10.1109/OJSP.2023.3337714.
37. Baloyi, E., Mahlasela, O., Siphambili, N. and Stegmann, M.: Social media as a strategic advantage during cyberwarfare: A systematic literature review. International Conference on Cyber Warfare and Security, vol. 19, no. 1, pp. 19–25, Mar. 2024. doi:10.34190/iccws.19.1.2036.
38. ENISA Threat Landscape 2023, Oct 2023.
39. Veerasamy, N., Badenhorst, D., Ntshangase, M., Baloyi, E., Siphambili, N., Mahlasela, O.: Unpacking AI security considerations," International Conference on Cyber Warfare and Security. vol. 19, no. 1, Mar. 2024. doi:10.34190/iccws.19.1.2104.
40. de Rancourt-Raymond, A. and Smaili, N.: The unethical use of deepfakes. Journal of Financial Crime, vol. 30, no. 4, pp. 1066–1077, May 2022. doi:10.1108/jfc-04-2022-0090.
41. Rini, R. and Cohen, L.: Deepfakes, deep harms. Journal of Ethics and Social Philosophy, vol. 22, no. 2, Jul. 2022. doi:10.26556/jesp.v22i2.1628.
42. Kietzmann, J., Lee, L.W., McCarthy, I. P. and Kietzmann, T. C.: Deepfakes: Trick or treat?. Business Horizons, vol. 63, no. 2, pp. 135–146, Mar. 2020. doi:10.1016/j.bushor.2019.11.006.
43. Olson. A.: "The Double-side of Deepfakes: Obstacles and Assets in the Fight Against Child Pornography," Georgia Law Review, vol. 56, no. 2 (2022)
44. Ahmed, S. Navigating the maze: Deepfakes, cognitive ability, and social media news skepticism. New Media &amp; Society, vol. 25, no. 5, pp. 1108–1129, Jun. 2021. doi:10.1177/14614448211019198.
45. Calamur, H.: Free Press Journal, https://www.freepressjournal.in/analysis/the-dark-side-of-artificial-intelligence-deepfakes-fake-news-and-ai-generated-images, last accessed 2024/03/20.
46. Thakur, A.: NDTV.com, https://www.ndtv.com/offbeat/ai-generated-images-of-barack-obama-and-angela-merkel-enjoying-vacation-on-a-beach-amazes-internet-3878630, last accessed 2024/03/30.
47. Chen. H and Magramo, K.: CNN, https://edition.cnn.com/2024/02/04/asia/deepfake-cfo-scam-hong-kong-intl-hnk/index.html, last accessed 2024/03/31.
48. Weng. L.: What are diffusion models? Lil'Log. https://lilianweng.github.io/posts/2021-07-11-diffusion-models/. (July, 2021).
49. Remya Revi, K., Vidya, K. R., & Wilscy, M.: Detection of deepfake images created using generative adversarial networks: A review. In Second International Conference on Networks and Advances in Computational Technologies: NetACT 19 (pp. 25-35). Springer International Publishing. (2021).

A Model for Predicting Continuance Intentions of mHealth with Community Health Workers in Malawi:

From User Expectations Perspective

First Author[1] [0000-1111-2222-3333] and Second Author[2] [1111-2222-3333-4444]

[1] Auth#
2Author#

Abstract This paper aims to model user expectations as predictors of continuance intentions of mHealth with community health workers (CHWs) in Malawi, a developing country context. The study extends the expectation confirmation model to include effort expectancy, and quality triads (system quality, information quality, and service quality). A survey questionnaire was used to collect data from 176 randomly sampled CHWs in three district health facilities in Malawi. Partial least squares method to structural equation modelling (PLS-SEM) was used to analyse data. The study found that effort expectancy, confirmation, satisfaction, and post-usage usefulness had a significant influence on CHWs' continuance usage intentions with mHealth in Malawi. However, the unexpected results were that quality triads did not yield positive effects on the continued usage intentions of CHWs with mHealth. This was contrary to the established IS extant literature. The study concludes by making recommendations for policy and research practice.

Keywords: User Expectations, mHealth, Satisfaction, Continuance intention, Cstock, Community Health Workers, Malawi, Developing Country, Africa

1 Introduction

The study uses Cstock as a case study to model user expectations as predictors of continuance intentions of mHealth with CHWs in Malawi. Cstock is a mHealth application used by community health workers in Malawi, to order and supply a stock of medicines from district health facilities to village clinics, delivered via short service message (SMS). The rapid growth in mobile phone subscriptions has encouraged Sub-Saharan African countries to use mobile health (mHealth) to solve a wide range of challenges in the health sector, such as health surveillance, supply chain management, and health education, among others (GSM, 2022).

mHealth is defined as the provision of health services and information through mobile technologies such as mobile phones and Personal Digital Assistants (Malanga, 2017; WHO, 2011). In the past decade, over 500 mHealth projects have been deployed in rural areas of Africa to improve and reduce the costs of patient monitoring, medical supply chain management, medical adherence, and healthcare worker communication

(Betjeman et al., 2013). Text messaging, voice calls, cameras, automated sensing video messaging, and the internet are some of the mobile technologies that are utilised in mHealth to improve access and quality of information services (Nyemba-Mudenda & Chigona, 2018).

Despite the proliferation of such mHealth initiatives, the majority of them have not moved beyond the pilot phase due to several challenges. These challenges vary from lack of sustainable financing, weak organisation structure, poor ICT infrastructure, lack of technical capacity, lack of interoperability and integration, and security and privacy concerns (Chipeta & Malanga; Malanga & Chigona, 2018; Mudenda & Chigona).

Prior studies indicate that the success of Information System (IS) use, such as mHealth is dependent on the user's continued use of the technology rather than on the initial acceptance (Battacherjee, 2001). IS continuance is defined as a repeated decision to use an information system after initial acceptance (Kim, Chan, & Chan, 2007). Whilst many factors can be attributed to this discontinued use of mHealth, users' expectations are suggested as critical factors for sustaining the continuance usage of mHealth. This assertion is also attested by prior studies that reported that a total of 33 critical factors were identified as determinants of IS success on various IS projects, and user expectations were ranked second (Gursel et. al., 2014; Peter, 2010). Hence, understanding about effects of user expectations on IS continuance research is an important step towards sustaining innovations like mHealth that users

1.1 User expectations and mHealth

User expectations are defined as a set of beliefs held by target users of an IS associated with IS and their performance of using the system (Szajna & Scamell, 1993). User expectations can reveal how users conceptualise the technology and how they expect it to benefit it (Olsson, 2014). This can help policymakers improve their users' satisfaction and continuance use of the new system (Linda, 2012). Despite the role that user expectations play in sustaining the adoption of IS projects, few studies have investigated how this phenomenon impacts IS continuance including mHealth.

Malawi, like other developing countries, has initiated several mHealth initiatives over the past five years being used by community health workers. Salient examples include Cstock, a mobile phone for health "Chipatala cha pa foni" among others (Malanga & Chigona, 2018). Prior studies on the Cstock and other mHealth applications in general so far have looked at the feasibility, implementation and acceptability of the technology (SC4CCM, 2018; Friedrichs et al., 2014; Shieshia et al, 2014). However, there is a dearth of literature regarding the influence of user expectations as predictors of continuance intentions of mHealth with community health workers in Malawi. Based on this limited evidence, this study was set out to fill this knowledge gap.

1.2 Research objectives

The main objective of the study was to investigate the potential user expectations as predictors of continuance intention of Cstock, mHealth application with CHWs in Malawi. Three specific objectives were posed: This was achieved by answering two specific objectives:

To identify potential user expectations as determinants of continuance intention of mHealth with CHWs in Malawi.

To examine the effects of user expectations on CHWs'post-usage usefulness with mHealth application.

To determine the effects of user expectations on the satisfaction of CHWs with mHealth application.

1.3 Cstock mHealth description

Cstock is a rapid Short Message Service (SMS) and web-based reporting and resupply system that is used by HSAs to report stock data about medical supplies through mobile phones. This mHealth application was designed by the Ministry of Health in partnership with Supply Chain for Community Case Management (SC4CCM) (SC4CCM, 2018). The system calculates HSAs to re-supply quantities and sends this information to the health facility staff. The health facility staff receive this information to pick and pack products for CHWS and notify them about the collection date and time (Shieshia et al, 2014). A web-based accessible dashboard is a simple, easy-to-use report, that shows stock levels, reporting rates, and alerts for central and district-level managers. The dashboard provides visibility of HSA logistics data to district and central-level managers. Recently, Cstock mHealth has been implemented in health facilities in 14 out of the 29 districts of the country (SC4CCM, 2018). Figure 1 illustrates how data and product flow in Cstock mHealth.

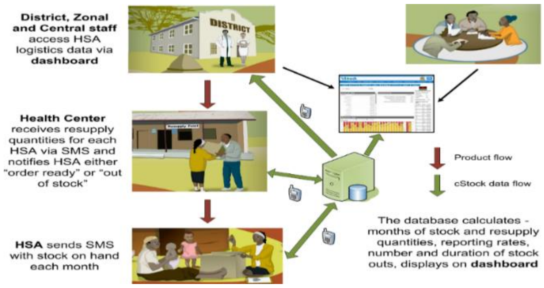

District, Zonal and Central staff access HSA logistics data via dashboard
DISTRICT
Health Center receives resupply quantities for each HSA via SMS and notifies HSA either "order ready" or "out of stock"
HSA sends SMS with stock on hand each month
Product flow
cStock data flow
The database calculates - months of stock and resupply quantities, reporting rates, number and duration of stock outs, displays on dashboard

Fig. 1. Cstock Data and Product Flow (SC4CCM, 2018)

2 Research model and hypotheses

The objective of this study was to model user expectations that can predict the continued usage intention of mHealth with the community health workers (CHWs). To conceptualise this study phenomenon, the researchers selected the expectation confirmation model (ECM) as a based model. The expectation confirmation model (ECM) is a model commonly used to study information System (IS) continued usage behaviour [29,28, 26,1].

ECM posits that an individual user's intention for continued use of information systems (IS) depends on three variables: the user's level of satisfaction with information systems, the extent of the user's confirmation of expectation, and the perceived usefulness, which later changed to post-usage usefulness [28, 1]. Based on the advantages of the ECM, this study adopted all four constructs of the model: (i) post-usage usefulness, (ii) confirmation, (iii) satisfaction, and (iv) continuance intention. Following the development of ECM, various studies have applied it to study post-adoption expectations in the mHealth domain. Recent salient examples include continuance behaviour in using e-Health services [30]; factors affecting users' continuance intention towards mHealth [31]; continuance of mHealth at the bottom of the pyramid [27]; continued usage intention of mHealth in a developing country [28]; and among others.

One of the strengths of ECM is that it is considered a parsimonious model and can be applied to different settings (Bhattacherjee et al., 2008). However, it has some shortcomings (D' Ambra et al., 2013). The model uses the post-usage usefulness as an aggregate variable for post-user expectations. Yet, post-usage usefulness alone is not adequate to capture all the post-expectation beliefs such as control, attitudinal, and object-based beliefs that an individual user may hold towards the IS. Second, ECM posits that pre-acceptance expectations are covered by satisfaction and confirmation constructs (Battacherjee, 2001). To address the weakness of the ECM, this study integrated quality triads (system quality, information quality, and service quality) from the IS success model (DeLone & McLean, 2016), and effort expectancy from the UTAUT model (Venkatesh et al., 2003). Thus, effort expectancy, system quality, information quality, and service quality were modelled as pre-usage/pre-acceptance expectations. The variables are briefly discussed:

2.1 Satisfaction

User satisfaction with technology such as mHealth denotes a positive aggregate feeling that develops by several dealings with the services (Osah &Kyobe, 2017; Kuo et al., 2009). Previous studies in consumer behaviour have found that satisfaction is a key determinant of customer's repurchase intentions (Oliver, 1980).

In the mHealth context, CHWs' satisfaction depends on what they want from Cstock, and what they realise after using the Cstock. This relationship has been validated by previous extant literature (Nie et al., 2023; Tian et al., 2022; Hossain et al., 2021; Kumar & Natarajan, 2020; Leung & Chen, 2019; Osah & Kyobe, 2017; Bhattacherjee et al., 2008; Bhattacherjee,2001). Thus, this study's first hypothesis is that:

H1: User’s high level of satisfaction with Cstock is positively associated with his or her continuance intention towards Cstock-mHealth application

2.2 Post-usage usefulness

Post-usage usefulness is regarded as a significant predictor of both user satisfaction and continued usage intentions of information system (IS) continuance (Bhattacherjee et al., 2008; Bhattacherjee, 2001). Thus, post-usage usefulness in ECM, represents the affirmative understanding of the performance or benefits accrued from using the technology artefact (Osah, 2015; Bhattacherjee, 2008). This relationship between post-usage usefulness and satisfaction has been validated by extant IS literature (Ayyoub et al.,2023; Nie et al., 2023; Osah & Kyobe, 2017; Ayanso et al., 2015; O'Brien, 2013; Bhattacherjee et al., 2008). Furthermore, the causal relationship between post-usage usefulness and continuance intention is validated in IS extant literature (Nie et al., 2023; Tian et al., 2022; Hossain et al., 2021; Kumar & Natarajan, 2020; Ayanso et al., 2015; Bhattacherjee, 2001). Therefore, this study’s second and third hypotheses are:

H2: User’s perceptions of post-usage usefulness are positively associated with his or her satisfaction with Cstock-mHealth application.

H3: User’s perceptions of post-usage usefulness are positively associated with his or her continuance intention towards Cstock-mHealth application.

2.3 Confirmation

In the consumer literature, confirmation entails the inconsistency between expectation and actual consumer experience with the product or service (Osah, 2017; Oliver, 1980). In IS literature, this view is reflected within the expectation confirmation model (ECM) (Bhattacherjee, 2001). This model posits that technology users’ confirmation of expectations asserts positive effects on their perception of usefulness and satisfaction with technology (Osah, 2015, Bhattacherjee, 2001). The extant IS literature that has validated the relationship between confirmation and post-usage usefulness include (e.g. Nie et al., 2023; Tian et al., 2022; Hossain et al., 2021; Kumar & Natarajan, 2020; Bhattacherjee, 2001). Similarly, the relationship between confirmation and satisfaction is evident in IS extant studies (Nie et al., 2023; Mtebe & Gallagher, 2022; Tian et al., 2022; Hossain et al., 2021; Fu et al., 2020; Baharum, & Jaafar, 2015; Bhattacherjee, 2001). Consistent with these findings, the fourth and fifth hypotheses for this study are:

H4: User’s high level of confirmation is positively associated with his or her post-usage usefulness with the Cstock-mHealth application.

Thus, the fifth hypothesis is that:

H5: The user’s high level of confirmation is positively associated with his or her satisfaction with the Cstock-mHealth application.

2.4 Effort expectancy

Effort expectancy (EF) is a variable reflected within the UTAUT model (Venkatesh et al., 2003). The theory posits that effort expectancy is the degree of ease of use associated with the use of the system. This implies that users of mHealth applications such as Cstock will accept the system if it is easy to use (Cavalcanti et al., 2022) and reject it if it is complex to use (Venkatesh et al.,2003).

Extant literature that has established the link between effort expectancy and usefulness is evident (e.g. Rezvani et al., 2022; Onaolapo & Oyewole, 2018; Sair &Danish, 2018). Similarly, previous IS studies have validated this relationship between effort expectancy and satisfaction (Elok, & Hidayati, 2021; Chao, 2019; Onalopo & Oyewole, 2018). Thus, the sixth and seventh hypotheses of this study were:

H6: The user's high level of perception of effort expectancy is positively associated with his or her post-usage usefulness with Cstock-mHealth application.

H7: The user's high level of perception of effort expectancy is positively associated with his or her satisfaction with the Cstock-mHealth application.

2.5 System quality

System quality is characterised by desirable features such as system flexibility, system reliability, ease of learning, response time, and among others (McLean & DeLone, 2016;2014; 2003). Likewise, in the mHealth context, the mHealth application should be characterised by these system quality attributes as posited by the ISS model (McLean & DeLone, 2016; 2014). Thus, if the mHealth application fails to exhibit these system quality attributes, it would be deemed of poor quality, this would jeopardise user experience (Osah, 2015).

Prior IS studies have validated the relationship between system quality and usefulness with technology. The findings revealed that system quality is a key determinant of system success and has a direct influence on the perception of usefulness (Alkhawaja et al., 2022; Haddad, 2018; Rui-Hsin & Lin, 2018) and satisfaction (Chen et al., 2024; Nurhayani et al., 2024) with technology. In light of the foregoing discussion and consistent with the findings, the researchers hypothesize that:

H8: The user's high level of perception of system quality is positively associated with his or her post-usage usefulness with the Cstock-mHealth application.

H9: The user's high level of perception of system quality is positively associated with his or her satisfaction with the Cstock-mHealth application

2.6 Information Quality

Information quality refers to the desirable characteristics of the system outputs (Chipeta & Malanga, 2022; McLean & DeLone, 2016). The information quality attributes include relevance, accuracy, completeness, timeliness, understandability, and conciseness, among others (McLean & DeLone, 2016). In the mHealth context, information delivered or received from the mHealth application that covers the desirable attributes of information quality will be well appreciated by the users and vice versa.

Previous IS studies have reported the relationship between information quality and satisfaction (Nurhayani et al., 2024; Alkhawaja et al., 2022; Rui-Hsin & Lin, 2018; Osah & Kyobe, 2017; Irma, 2014; Osah, 2015). In addition, the relationship between information quality and post-usage usefulness is confirmed by prior IS extant studies (Machdar, 2016; Irma, 2014; Wijayanto, 2008; Seddon, 1997). Thus, consistent with the findings from extant studies, it was hypothesized that:

H10: User's high level of perception of information quality is positively associated with his or her satisfaction with Cstock-mHealth application

H11: User's high level of perception of information quality is positively associated with his or her post-usage usefulness with the Cstock-mHealth application.

2.7 Service Quality

Service quality refers to the quality of support that information technology (IT) users receive from the information organisation and IT staff (McLean & DeLeon, 2016; 2003). The service quality characteristics include reliability, technical competency and empathy of technical IT staff, assurance, personalisation, accuracy, and responsiveness (Osah&Kyobe,2017; McLean & DeLone, 2016).

With the mHealth application, where users' experiences low responsiveness and slow connections and there is no competent and empathetic IT staff, this is likely to upset the users' experiences with negative perceptions of usefulness while using mHealth application and vice-versa. Prior extant studies have a positive relationship between service quality and post-usage usefulness (eg. Shah & Attiq, 2016; Calisir et al., 2014; Kim & Lee, 2014; Wang & Xiao, 2009). Furthermore, there also evidence for relationship between service quality and satisfaction is reported in prior IS extant studies (Chen et al., 2024; Nurhayani et al., 2024; Alkhawaja et al., 2022; Rui-Hsin & Lin, 2018; Osah & Kyobe, 2017; Osah, 2015). Accordingly, it was hypothesized that:

H12: *User's high level of perception of service quality is positively associated with his or her post-usage usefulness with the Cstock-mHealth application.*

H13: *User's high level of perception of service quality is positively associated with his or her satisfaction with the Cstock-mHealth application.*

Figure 2, provides the hypothesized research model for this study.

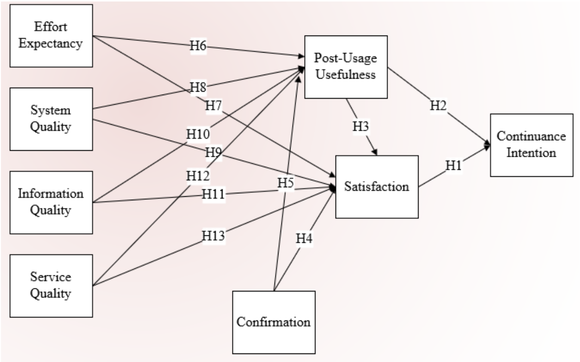


Fig. 2. Research model for this study

3 Methodology

3.1 Research design

This study adopted a cross-sectional survey design, employing a quantitative method (Saunders et al. 2016).This survey design deals with the opinions of individuals that are collected at one point in time. The reasons for employing survey design were that it is popular and allows the collection of a large amount of data from a sizeable population (Creswell, 2014). This study adopted a quantitative research method. This study utilised a print survey questionnaire as a data collection instrument. The survey questionnaires consisted of 30 closed-ended standardised questions measuring nine variables adapted from extant theories and literature. The questions were subjected to a 7-point Likert scale (1=strongly disagree to 7=strongly agree). This enhanced the validity and reliability of the research instrument items. About 280 survey questionnaires were distributed randomly to health surveillance assistants (HSAs) who were users of Cstock in three district health facilities in Malawi.

The questions from the constructs were adapted from the previous validated studies as follows: continuance intention (CONT) (Bhattacherjee, et al., 200; Bhattacherjee, 2001); post-usage usefulness (PUU) (Alkhawaja et al., 2022; Haddad, 2018); satisfaction (SATIS) (Nie et al., 2023; Osah & Kyobe, 2017); confirmation (CONF)(Tian et al., 2022; Bhattacherjee et al.,2008);

Effort expectancy (EEXP) (Rezvani et al., 2022; Onaolapo & Oyewole, 2018); system quality (SYSQ) (Fu et al., 2023; McLean & DeLone, 2016); information quality (INFQ) (Nurhayani et al., 2024; McLean & DeLone, 2016); and service quality (SERVQ) (Chen et al., 2024; Shah & Attiq, 2016).

3.2 Ethical clearance

The researchers obtained research ethical approval from the University of BBUX (Pseudonym), with reference HREC REF: 553/2019. Furthermore, permission to collect data was sought from the three district health facilities that were targeted (Nkhata-Bay, Chitipa, and Rumphi). The participation was voluntary such that each participant was free to take part or withdraw in the study at any time. Consent forms were also given to each respondent to sign his/her agreement or willingness to participate in the study. Creswell, 2014).

3.3 Data collection and analysis

The study utilised a partial least square method of structural equation modelling (PLS-SEM) as a method of data analysis. PLS-SEM examines both the measurement and structural models simultaneously, consequently assessing factor analysis and hypothesis testing concurrently (Hair et al., 2014). In addition, PLS-SEM demands on smaller sample size than CB-SEM (Hair et al., 2017). The sample size for this study is 176 CHWs/HSAs more than 157 sample size, which is enough to satisfy the recommendation of PLS-SEM (Cohen, 1992). Data collection took place between July 2020 to May 2021. Of 280 survey questionnaires that were randomly distributed to CWHS/HSAs in the three district health facilities (Chitipa=90, Rumphi=120, and Nkhatabay=70), about 227 were successfully returned for data analysis, achieving a response rate of 81.07%. SmartPLS version 3.3.9 software tool was used to analyse the data.

3.4 Data Screening

To screen data for quality, data was submitted to Haman's one-factor method in IBM SPSS version 25. The results showed that the first factor accounted for 12.608%, while the last factor was 5.058%. The results also indicated that all 8 factors that were submitted to the Principal Component Analysis (PCA) accounted for 62.492%. Thus, a single factor was not accountable for more than 50% of the variance in the data set, indicating the absence of CMB (Barns & Barns, 2008). In addition, data normality was assessed by using the D'Agostino Skewness and Ansombe-Glynn Kurtois tests (Hossan et al., 2020; Hair et al., 2017). Skewness is used to describe the balance of distribution, and the recommended Z-scores of -2+2 were employed as a threshold (Hair et al., 2014). Azzalini (2005). Likewise, Kurtois refers to the peakedness or flatness of the distribution compared to the normal distribution, and threshold values of -3 to +3 were adopted in this study (Azzalini, 2005). Thus, after screening the data, 51 cases were removed and only 176 cases were returned for data analysis.

4 Empirical Analysis of Findings

4.1 Demographic profile of respondents

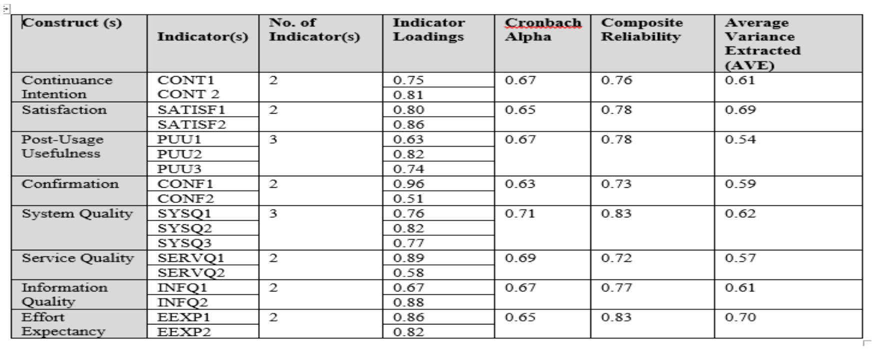
The first part of the survey sought to gather the demographic profile of respondents. The results showed that more males (60.08% or 107) participated in the study than females (39.2% or 69). More respondents between the 26-30 age range (37.5% or 66) participated in the survey, followed by those aged between 36-40 (36.4% or 64), while only 15.3% (27) were aged above 41 and above. With regards to education, the majority of respondents were holders of Malawi School Certificate of Education (MSCE) (67.0% or 118) and 1.1% (2) were holders of degrees. In terms of district health facilities, 38.1% (67) of respondents were from Chitipa, 33.0% (58) were from Nkhata-bay, and 29.0% (51). Besides, about work experience, the majority (48.9% or 86) of respondents who participated in the survey had 11-15 years of work experience followed by 26.7% (67) of those who had 6-10 years of work experience. The study also revealed that 71.6% (126) were using Cstock mHealth monthly. Moreover, the study also found that 48.3% (81) of respondents were using featured phones to access Cstock mHealth, followed by 43.1% (76) of those who were using basic phones to access the Cstock, mHealth application.

4.2 Assessment of the Measurement Model

The first step involves the evaluation the of reliability and validity of constructs in the conceptualized model (Hair et al., 2021; Knock, 2014). Using SMART-PLS 3.8, this study followed four steps to evaluate the reliability and validity of the research model: (i) Indicator reliability: Indicator loadings of 0.50 were adopted (Hair et al.,2017);(ii) Internal consistency reliability: Crobach alpha values of 0.60 and composite reliability of 0.70 were adopted respectively (Chin, 2010; Hamid, Sami & Sadek, 2017); and (iii) Convergent validity: the Average Variance Extracted (AVE) with 0.50 was examined (Sarstedt, Ringle & Hair, 2017; Fornell-Lacker, 1981).

In this study, a path algorithm through SMART-PLS results was performed. The results show that all the values that met the quality criteria for establishing reliability and convergent validity were returned, while those that did not meet the criteria were dropped off or removed. The summary of results for examining reliability and convergent validity are presented in Table 2.

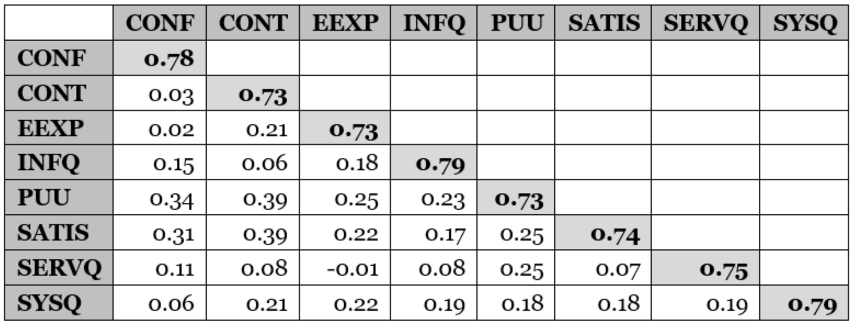
Table 1. Results for testing reliability and convergent validity



Construct (s)	Indicator(s)	No. of Indicator(s)	Indicator Loadings	Cronbach Alpha	Composite Reliability	Average Variance Extracted (AVE)
Continuance Intention	CONT1 CONT 2	2	0.75	0.67	0.76	0.61
			0.81			
Satisfaction	SATISF1	2	0.80	0.65	0.78	0.69
	SATISF2		0.86			
Post-Usage Usefulness	PUU1	3	0.63	0.67	0.78	0.54
	PUU2		0.82			
	PUU3		0.74			
Confirmation	CONF1	2	0.96	0.63	0.73	0.59
	CONF2		0.51			
System Quality	SYSQ1	3	0.76	0.71	0.83	0.62
	SYSQ2		0.82			
	SYSQ3		0.77			
Service Quality	SERVQ1	2	0.89	0.69	0.72	0.57
	SERVQ2		0.58			
Information Quality	INFQ1	2	0.67	0.67	0.77	0.61
	INFQ2		0.88			
Effort Expectancy	EEXP1	2	0.86	0.65	0.83	0.70
	EEXP2		0.82			

Another method that was employed to examine the validity of the study is to assess the discriminant validity. This measures the extent to which a construct is empirically distinct from other constructs both in terms of how much it correlates with other constructs and distinctly the indicators represent only the single construct (Hair et al, 2017; Sarstedt, Ringle & Hair, 2017). To establish discriminant validity, this study examined the Fornell and Lacker criterion (Fornell-Lacker, 1981) and cross-loadings (Henseler et al., 2015). To establish discriminant validity via the Fornell-Lacker Criterion, the square root of each construct's AVE should be greater than its highest correlation with any other construct (Hair et al., 2014). Thus, this study examined the Fornell-Lacker criterion using SMART-PLS, and the results are presented in Table 3. The results indicate that the square roots of each construct's AVE are greater than its highest correlation with any other construct, thereby indicating the presence of discriminant validity.

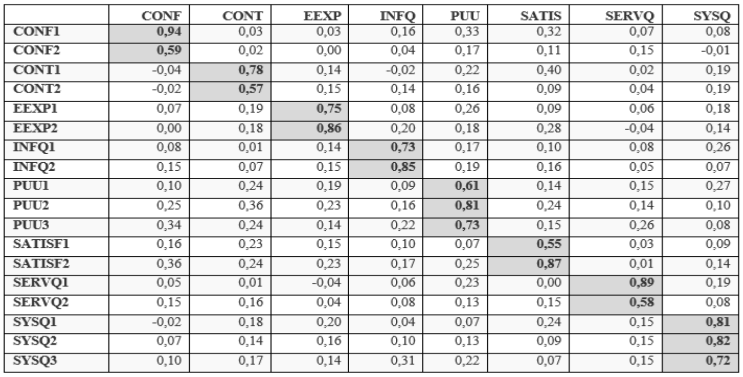
Table 2. Results of Fornel-Lacker Criterion



	CONF	CONT	EEXP	INFQ	PUU	SATIS	SERVQ	SYSQ
CONF	**0.78**							
CONT	0.03	**0.73**						
EEXP	0.02	0.21	**0.73**					
INFQ	0.15	0.06	0.18	**0.79**				
PUU	0.34	0.39	0.25	0.23	**0.73**			
SATIS	0.31	0.39	0.22	0.17	0.25	**0.74**		
SERVQ	0.11	0.08	-0.01	0.08	0.25	0.07	**0.75**	
SYSQ	0.06	0.21	0.22	0.19	0.18	0.18	0.19	**0.79**

Through cross-loadings, discriminant validity is shown when each measurement item correlates weakly with all other constructs except for the one with which it is theoretically associated. To establish discriminant validity in PLS-SEM analysis using cross-loadings, Chin (1998) suggests that each indicator loading should be greater than all of its cross-loadings. Likewise, as indicated in Table 4, the results show that all indicators have greater loadings than all of their cross-loadings, implying the presence of discriminant validity.

Table 3. Results of Cross-Loadings



	CONF	CONT	EEXP	INFQ	PUU	SATIS	SERVQ	SYSQ
CONF1	**0,94**	0,03	0,03	0,16	0,33	0,32	0,07	0,08
CONF2	**0,59**	0,02	0,00	0,04	0,17	0,11	0,15	-0,01
CONT1	-0,04	**0,78**	0,14	-0,02	0,22	0,40	0,02	0,19
CONT2	-0,02	**0,57**	0,15	0,14	0,16	0,09	0,04	0,19
EEXP1	0,07	0,19	**0,75**	0,08	0,26	0,09	0,06	0,18
EEXP2	0,00	0,18	**0,86**	0,20	0,18	0,28	-0,04	0,14
INFQ1	0,08	0,01	0,14	**0,73**	0,17	0,10	0,08	0,26
INFQ2	0,15	0,07	0,15	**0,85**	0,19	0,16	0,05	0,07
PUU1	0,10	0,24	0,19	0,09	**0,61**	0,14	0,15	0,27
PUU2	0,25	0,36	0,23	0,16	**0,81**	0,24	0,14	0,10
PUU3	0,34	0,24	0,14	0,22	**0,73**	0,15	0,26	0,08
SATISF1	0,16	0,23	0,15	0,10	0,07	**0,55**	0,03	0,09
SATISF2	0,36	0,24	0,23	0,17	0,25	**0,87**	0,01	0,14
SERVQ1	0,05	0,01	-0,04	0,06	0,23	0,00	**0,89**	0,19
SERVQ2	0,15	0,16	0,04	0,08	0,13	0,15	**0,58**	0,08
SYSQ1	-0,02	0,18	0,20	0,04	0,07	0,24	0,15	**0,81**
SYSQ2	0,07	0,14	0,16	0,10	0,13	0,09	0,15	**0,82**
SYSQ3	0,10	0,17	0,14	0,31	0,22	0,07	0,15	**0,72**

4.3 Assessment of the Structural Model

The structural path model in PLS-SEM was assessed by examining collinearity, explanatory power of the structural model and the path coefficient (Sarstedt et al., 2019). Collinearity issues arise when there is a correlation or association between two or more predictor variables in a structural path model (SAGE, 2019; Hair et al., 2017). This study adopted measures of tolerance and variance inflation factor (VIF) values. As a threshold, where each predictor construct or variable's tolerance and VIF should be above 0.20 but less than 5 respectively (Hair et al., (2014). To compute tolerance and

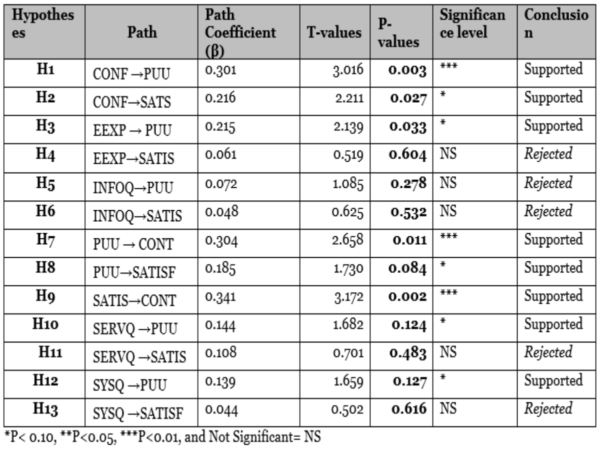
VIF values, each set of predictor variables and the latent scores from the default report in SMART-PLS were copied (Osah, 2017). Table 5, indicates that all the tolerance and VIF values are above 0.20 and less than 5, implying that collinearity was not an issue in the study.

Table 4. Results for testing multicollinearity

NO	**Construct**	**Tolerance**	**VIF**
1	Confirmation	0.87	1.15
2	Effort Expectancy	0.85	1.18
3	Information Quality	0.93	1.07
4	Post-Usage Usefulness	0.85	1.17
5	Satisfaction	0.86	1.16
6	Service Quality	0.98	1.02
7	System Quality (SYSQ)	0.62	1.61

To examine the structural model of the path coefficients, a bootstrapping method was run in SMART-PLS to obtain the t-values, p-values and standard errors to determine the statistical significance of the proposed hypotheses. To achieve this, a critical value of two-tailed tests was adopted as a rules of thumb postulated in IS literature. Table 5 shows that 8 hypotheses had significant path relationships (HI, H2, H3, H7, H8, H9, H10, and H12) while 5 hypotheses were rejected (H4, H5, H6, H11, H13).

Table 5. Results of examining path coefficients in the structural model



Hypotheses	**Path**	**Path Coefficient (β)**	**T-values**	**p-values**	**Significance level**	**Conclusion**
H1	CONF →PUU	0.301	3.016	**0.003**	***	Supported
H2	CONF→SATS	0.216	2.211	**0.027**	*	Supported
H3	EEXP → PUU	0.215	2.139	**0.033**	*	Supported
H4	EEXP→SATIS	0.061	0.519	**0.604**	NS	*Rejected*
H5	INFOQ→PUU	0.072	1.085	**0.278**	NS	*Rejected*
H6	INFOQ→SATIS	0.048	0.625	**0.532**	NS	*Rejected*
H7	PUU → CONT	0.304	2.658	**0.011**	***	Supported
H8	PUU→SATISF	0.185	1.730	**0.084**	*	Supported
H9	SATIS→CONT	0.341	3.172	**0.002**	***	Supported
H10	SERVQ →PUU	0.144	1.682	**0.124**	*	Supported
H11	SERVQ →SATIS	0.108	0.701	**0.483**	NS	*Rejected*
H12	SYSQ →PUU	0.139	1.659	**0.127**	*	Supported
H13	SYSQ →SATISF	0.044	0.502	**0.616**	NS	*Rejected*

*P< 0.10, **P<0.05, ***P<0.01, and Not Significant= NS

This study also examined the coefficient of determination (R^2 value) of the structural path model. It is a measure of the predictive power of the model and calculated as the squared correlation between a specific endogenous construct's actual and predicted values (Hair et al. 2016). The rules of thumb acceptable for R values depend on the complexity of the research and discipline. For instance, in marketing literature, R^2 values of 0.75, 0.50 or 0.25 for criterion variables are considered as substantial, moderate or weak respectively (Hair et al.,2011)). In this study, through the bootstrapping method in SMARTPLS, an R^2 value of 0.275 was obtained from the criterion variable (criterion variable), and thus can be considered higher since this study was exploratory. Figure 3 illustrates the outer loadings, path coefficients and the R^2 value explained by the conceptualised structural path model.

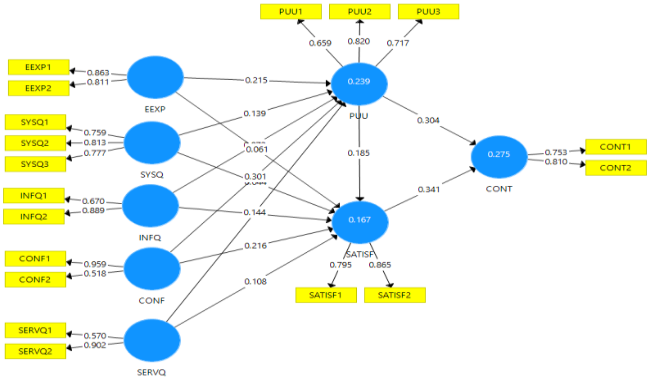


Fig. 3. Results of Coefficient of determination (R^2)

5 Discussion of Findings

First, the post-usefulness (PUU), satisfaction (SATIS), confirmation (CONF), and continuance intention (CONT) were drawn from ECM (Bhattacherjee et al., 2008). Three of the four were predictor variables and continuance intention was a criterion variable. Five path relationships were hypothesised. The study found that all the five hypotheses had significant path relationships. Post-usage usefulness represents the affirmative understanding of the performance or benefits accrued from using the technology artefact (Osah, 2015; Bhattacherjee, 2008). The findings imply that CHWs valued the accrued benefits realized from using the Cstock, mHealth application. Previous studies have revealed that post-usage usefulness has a positive effect on continuance intention (Ayyoub et al.,2023; Nie et al., 2023).

Besides, the study found that CHWs were satisfied with services delivered by the Cstock, mHealth application. Satisfaction emerged as the strongest predictor of continuance intention with Cstock, mHealth application. This is also confirmed by ECM as postulated by Bhattacherjee (2008). This positive relationship entails that the sampled users are content with the functionalities offered by Cstock. The result was also consistent with the IS extant literature (Nie et al., 2023; Tian et al., 2022; Hossain et al., 2021; Kumar & Natarajan, 2020).

Moreover, the study reported a positive significant relationship between post-usage usefulness and satisfaction. The findings were consistent with previous literature (Ayyoub et al.,2023; Nie et al., 2023; Osah & Kyobe, 2017) The significance of this path relationships implies that CHWs recognize that their continued use of Cstock, mHealth application depends on the level of satisfaction following each usage of the system. It also signifies that users may get updated with new information, which further determines the level of users' satisfaction with technology at any given time (Osah, 2015). Thus, paying more attention to the benefits and increasing the level of satisfaction of CHWs will motivate their intentions to continue using Cstock, mHealth application despite the available alternatives to mHealth technologies.

Second, system quality (SYSQ), information quality (INFQ) and service quality (SERVQ) were drawn from the updated IS success model (DeLone & McLean, 2016). Six hypotheses were conceptualised as pre-acceptance expectations to predict PUU and SATIS. The findings emerged that only two hypotheses had a significant influence on post-usage usefulness.

System quality is characterized by desirable features such as system flexibility, system reliability, ease of learning, response time, and among others (McLean & DeLone, 2016;2014; 2003). On the other hand, service quality refers to the quality of support that information technology (IT) users receive from the information organisation and IT staff (McLean & DeLeon, 2016; 2003). The significance of the two hypotheses meant that the surveyed sampled CHWs recognized the importance of system quality (e.g. learnability, understandability) and service quality features (e.g. assurance, responsiveness) portrayed by Cstock, mHealth could impact positively their cumulated future benefits of using the system. The findings were consistent with previous studies (Shah & Attiq, 2016; Calisir et al., 2014; Alkhawaja et al., 2022; Haddad, 2018).

The insignificance of the other four hypotheses implied that the surveyed Cstock users developed their level of satisfaction with the system independent of system quality, information quality, and service quality features or characteristics. However, the findings were inconsistent previous studies, where positive relationships were reported between information quality and post-usefulness (Machdar, 2016; Irma, 2014; Wijayanto, 2008) and satisfaction (Chen et al., 2024; Nurhayani et al., 2024); system quality and satisfaction (Chen et al., 2024; Nurhayani et al., 2024); and service quality on satisfaction (Chen et al., 2024; Nurhayani et al., 2024). To this end, the rejection of these hypotheses also implies that CHWs' perceived information quality and service quality expectations were not met to instigate satisfaction and accrued benefits while using the Cstock, mHealth application.

Third, Effort expectancy (EEXP) was the only construct drawn from the UTAUT model, as a predictor variable (Venkatesh et al., 2003). Two path relationships were hypothesised as pre-acceptance expectations to predict PUU and SATIS. However, only the path relationship between EEXP and PUU was found significant. Effort expectancy is the degree of ease of use associated with the use of the system. This implies that users will accept Cstock, an mHealth application, if it is easy to use and reject it if it is complex (Cavalcanti et al., 2022; Venkatesh et al.,2003).

Thus, the positive relationship between effort expectancy and post-usage usefulness means that the surveyed sampled CHWs viewed effort expectancy features had a positive impact on the post-usefulness of Cstock, mHealth application. The findings were confirmed by previous studies (Rezvani et al., 2022; Onaolapo & Oyewole, 2018). However, the insignificance of effort expectancy on satisfaction, implies that users did not consider this variable playing a significant impact on the level of satisfaction and continuance intentions to use the Cstock, mHealth application. Thus, the Cstock, mHealth application vendor/provider should improve the effort expectancy characteristics of Cstock so that users are satisfied with the system.

6 Conclusion

This paper was set out to investigate the influence of user expectations on continuance intentions to use Cstock. This a mHealth application is used by CHWs via short message services (SMS) to order and supply medical supplies in village clinics in Malawi. The study has found that effort expectancy, system quality, trust and confirmation have a positive influence on continuance intentions towards Cstock mediated through post-usage usefulness and satisfaction. The findings have reinforced and also contested conventional ways of explaining user expectations and continuance intention. The conventional views that user's continuance intention towards technology is dependent on user's level of satisfaction, the level of user's confirmation of expectations, and post-usage usefulness (post-adoption expectations) were validated (Osah & Kyobe, 2017; Bhattacherjee et al., 2008; Bhattacherjee, 2001). However, the study found that system quality, information quality, and service quality did not have a positive influence on user satisfaction with mHealth. This implies that the surveyed users of Cstock, mHealth application in Malawi possess different expectation beliefs that may influence their level of satisfaction with technology contrary to the extant literature that views quality triads as determinants of user satisfaction with technology and its continuance intention (DeLone and McLean, 2016: Bhattacherjee et al., 2008; Bhattacherjee, 2001).

6.1 Theoretical implications

The empirical findings from this study demonstrate the relevance of the proposed research model, which is based on ECM with some additional expectation factors drawn from other theoretical models from extant IS pre-acceptance literature. The study also provides evidence of the relationships between user pre-acceptance expectations and post-adoption expectations on continuance intention towards mHealth.

Equally important, by integrating variables from pre-acceptance models to ECM, this study has refuted earlier studies (Bhattacherjee et al., 2008, 2001) that it was impossible to integrate both pre-acceptance and continuance theories in a single study due to variant time-bound.

6.2 Practical implications

These research findings may inform policymakers and research practitioners on the importance of taking into account user's expectations when deploying mHealth technologies in Malawi and beyond. The study also noted that user expectations that influence continuance usage intentions of mHealth among users in developing countries differ from those from the developed world. For, instance, there is established knowledge wisdom that quality triads have a positive influence on user satisfaction (McLean & DeLone, 2016). However, in the Malawian context, these quality triads were not significant, implying that the level of satisfaction users of Cstock, and mHealth application was not impacted by quality expectation features portrayed by the system. Furthermore, the study found that system quality, information quality and service quality did not have positive effects on the satisfaction of CHWs with the Cstock, mHealth application. Thus, it is recommended that Cstock, mHealth application provider/vendor should improve the quality features to ensure CHWs' high level of satisfaction with Cstock, mHealth and their ultimate continued usage behaviour.

6.3 Limitations and future work

Despite the theoretical and practical contributions, this study has also some limitations that warrant areas for future research. This study was conducted as a cross-sectional survey design targeting only CHWs of Cstock, mHealth application in three district health facilities of Malawi. Therefore, future studies must utilise longitudinal approach and replicate the study in other district health facilities in Malawi, where the Cstock, mHealth application is in use. Second, the study employed a quantitative research method, which also has its weakness. It is suggested that future studies should utilise mixed research methods to gain in-depth understanding of the study phenomenon. In the end, this will warrant generalization of the findings.

References

1. Akter, S., Ambra, J. D., Ray, P., & Hani, U. (2013). Modelling the impact of mHealth service quality on satisfaction continuance and quality of life, 3001. https://doi.org/10.1080/0144929X.2012.745606
2. Alruwaie, M., EL-Haddadeh, R., & Weerakkody, V. (2012). A framework for evaluating citizens ' outcome expectations and satisfactions toward continued intention to use e-Government services. Doctoral Symposium, Brunei Business School, (27th & 28th March), 1–12.
3. Ary, D. Jacobs, L.C., Sorensen, C.K. & Walter, D.A. (2014). Introduction to Research in Education. 9th ed. London: Wadsworth
4. Bhattacherjee, A. (2001). Understanding Information Systems Continuance: An Expectation Confirmation Model. MIS Quarterly, 25(3), 351–370. https://doi.org/10.2307/3250921
5. Brown, S. A., Venkatesh, V., Kuruzovich, J., & Massey, A. P. (2008). Expectation confirmation: An examination of three competing models. Organizational Behavior and Human Decision Processes, 105(1), 52–66. https://doi.org/10.1016/j.obhdp.2006.09.008

6. Chin, W. W., & Newsted, P. R. (1999). Structural equation modelling analysis with small samples using partial least squares
7. Creswell, J.W. (2014). Qualitative, Quantitative and Mixed Methods Approaches. 4th ed. London: SAGE Publication.
8. DeLone, W. H., & McLean, E. R. (2004). Measuring e-Commerce Success : Applying the DeLone & McLean Information Systems Success Model. International Journal of Electronic Commerce, 9(1), 31–47. https://doi.org/10.1080/10864415.2004.11044317
9. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of marketing research, 39-50.
10. Gagnon, M. P., Ngangue, P., Payne-Gagnon, J., & Desmartis, M. (2016). M-Health adoption by healthcare professionals: A systematic review. Journal of the American Medical Informatics Association, 23(1), 212–220. https://doi.org/10.1093/jamia/ocv052
11. Gilani,M.S., Iranmanesh,M., Nikbin, D., & Zailani,S. (2016). EMR Continuance usage intention of healthcare professionals. Health informatics and social care. http://dx.doi.org/10.3109/17538157.2016.1160245
12. Gürsel, G. (2015). Perceived Importance of User Expectations from Healthcare Information Systems, 84–86. https://doi.org/10.4018/978-1-4666-6316-9.ch005
13. Gürsel, G. (2016). User expectations: Nurses' perspective. Studies in Health Technology and Informatics, 225, 897–898. https://doi.org/10.3233/978-1-61499-658-3-897
14. Hair, J.F., Black, W.C., Babin, B.J., & Anderson, R.E. (2014). Multivariate Data Analysis (7th Ed).Harlow: FT/Prentice Hall
15. John, O.P., & Benet-Martinez, V. (2000). Measurement: Reliability, Construct Validation, and Scale
16. Kadzandira, J.M.,& Chilowa, W(2001) The Role of Health Surveillance Assistants (HSAs) in the Delivery of Health Services and Immunisation in Malawi
17. Larsen-Cooper, E., Bancroft, E., Rajagopal, S., O'Toole, M., & Levin, A. (2016). Scale Matters: A Cost-Outcome Analysis of an m-Health Intervention in Malawi. Telemedicine and E-Health, 22(4), 317–324. https://doi.org/10.1089/tmj.2015.0060
18. Leedy, P. & Ormrod, J. (2001). Practical research: Planning and design (7th Ed.). Upper Saddle River, NJ: Merrill Prentice Hall. Thousand Oaks: SAGE Publications.
19. Lemaire, J. (2011). Scaling Up Mobile Health: Elements Necessary for the Successful Scale Up of mHealth in Developing Countries.White paper commissioned by Advanced Development for Africa.Retrieved April 6, 2017, from https://www.k4health.org/sites/default/files/ADA_mHealth%20White%20Paper.pdf
20. Linda, S. L. (2012). Managing User Expectations in Information Systems Development. International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering, 6(12), 3710–3714. Retrieved from https://waset.org/Publication/managing-user-expectations-in-information-systems-development/14658
21. Malanga, D. F., & Chigona, W. (2018). Mobile Health Initiatives in Malawi. Understanding Impact, Funding Model and Challenges. International Journal of Privacy and Health Information Management, 6(1), 49–60. https://doi.org/10.4018/IJPHIM.2018010104
22. Mburu, S. (2017). Application of Structural Equation Modelling to Predict Acceptance and Use of mHealth Interventions at the Design Stage, 11(2), 1–17.
23. Miller, H. (2017). Managing customer expectations, 530(November). https://doi.org/10.1201/1078/43191.17.2.20000301/31233.12
24. Min, Q. (2007). An Extended Expectation Confirmation Model for Information Systems Continuance, 3879–3882.
25. Nyasulu C. & Chawinga, W.D (2018) The role of information and communication technologies in the delivery of health services in rural communities: Experiences from Malawi. SA

Journal of Information Management https://www.researchgate.net/deref/https%3A%2F%2Fdoi.org%2F10.4102%2Fsajim.v20i1.888

26. Ojasalo, J. (2006). Research and concepts Managing customer expectations in professional services. Managing Service Quality: An International Journal, Vol. 11 Issue: 3, pp.200-212, https://doi.org/10.1108/09604520110391379
27. Oliver, R. L. (1980). A cognitive model of the antecedents and consequences of satisfaction decisions. Journal of Marketing Research, 460-469
28. Petter, S. (2008). Managing user expectations on software projects: Lessons from the trenches. International Journal of Project Management, 26(7), 700–712. https://doi.org/10.1016/j.ijproman.2008.05.014
29. Randy Ryker,I., Nath, R., & Henson, J. (1997). Determinants of Computer user expectations and their relationship satisfaction with user satisfaction: An empirical Study, 33(4), 529–537.
30. Saunders, M., Lewis, P. & Thornhill, A. (2012). Research Methods for Business Students. 6th ed. London: Apprentice Hall
31. SC4CCM (2018). Emerging Lessons of Cstock. http://sc4ccm.jsi.com/emerging-lessons/cstock/
32. Shieshia, M., Noel, M., Andersson, S., Felling, B., Alva, S., Agarwal, S., Lefevre, A., Misomali, A., Chimphanga, B., Nsona, H., … Chandani, Y. (2014). Strengthening community health supply chain performance through an integrated approach: Using mHealth technology and multilevel teams in Malawi. Journal of Global Health, 4(2), 020406.
33. Straub, D., Boudreau, M. C., & Gefen, D. (2004). Validation guidelines for IS positivist research. The Construction. In H. Reis & Judd, C. (Eds.), Handbook of Research Methods in Social Psychology, 339-369. New York: Cambridge University Press.
34. Szajna, B., & Scamell, R. W. (1993). The Effects of Information System User Expectations on Their Performance and Perceptions. MIS Quarterly, 17(4), 493–516. https://doi.org/10.2307/249589
35. Taylor, S. and Todd, P. A. (1995). Understanding information technology usage: A test of competing models. Information systems research, 6(2):144–176.
36. Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User Acceptance of Information

THE FOUR PROCESSES OF AN EFFECTIVE CYBER SECURITY POLICY

Iyaloo Haitula-Waiganjo[1] Jude Osakwe[2] Ambrose Azeta[3]

[1,2,3] Namibia University of Science and Technology
Windhoek
13 Jackson Kaujeua St, Windhoek

Abstract. This article explores the process of cybersecurity policy formulation, implementation, and modification, emphasising the critical role of policy compliance in fortifying organizational digital defences. Drawing insights from various literature sources, the article highlights the multifaceted nature of cybersecurity policies, encompassing technological, procedural, and human-centric elements. The policymaking steps, including formulation, implementation, modification, and compliance, are described, underscoring the importance of tailoring policies to unique organizational cyber platforms. The study identifies and elaborates on essential cybersecurity policies, such as privacy, email security, net- work security, Wi-Fi usage, physical security, password management, and incident response. The article also introduces Lubua and Pretorius's cyber-security policy framework, illustrating seven key entries for comprehensive policy development. Furthermore, it stresses the ongoing need for policy compliance as a cornerstone for effective cybersecurity within organizations, involving both technical and non-technical solutions. The dynamic nature of technology and the continuous evolution of cyber threats necessitate periodic reviews and modifications to cybersecurity policies. The iterative process of cybersecurity policy development, implementation, compliance, and modification establish a robust frame- work, safeguarding digital infrastructure and enabling effective responses to evolving cyber challenges.

Keywords: Cybersecurity, Cybersecurity Policy, Cybersecurity policy formulation, Cybersecurity policy Compliance.

1 Introduction

The information security or cybersecurity policy encourages proper conduct among staff members by outlining duties and expectations for adhering to the rules and regulations of said policies [1]. These policies serve as the blueprint guiding an organisation's actions to secure its entire cyberspace. They outline a comprehensive roadmap, encompassing technological, procedural, and human-centric elements aimed at fortifying digital assets and mitigating risks. On the other hand, policies within the cybersecurity domain constitute a set of legislations, programs, and actions formulated by a governing body within an organisation. These policies serve as regulatory frameworks, dictating the protocols and standards to be followed to safeguard information and physical assets from potential threats [2][3].

Employees are usually the target audience for cybersecurity policies, as noted by [1] and they should always be taken into account when developing policies. Employees who interact with technology must be involved in cybersecurity governance, strategies, and policies because they are essential to strengthening an organisation's cyber defences, mitigating vulnerabilities, and proactively managing potential threats and attacks in the digital sphere. Furthermore, [4] draw the conclusion that having technological solutions for cybersecurity is not as crucial as an organisation establishing a set of information security policies and procedures.

Therefore, the purpose of this article is to give readers a better understanding of the cybersecurity policymaking process and its importance in protecting organizations against cyber threats. It seeks to explore the various steps involved in formulating, implementing, and ensuring compliance with cybersecurity policies. Additionally, the article aims to highlight the importance of tailoring policies to address the unique cyber platforms of organizations and fostering a culture of security and resilience. Through insights drawn from scholarly literature, the article aims to equip readers with knowledge and insights to navigate the complexities of cybersecurity governance effectively. Ultimately, the goal is to help organizations develop robust cybersecurity frameworks that can adapt to evolving threats and technologies, thereby enhancing their cybersecurity posture and mitigating risks. The following section will discuss the procedural steps in policymaking outlined by [6].

2 Cybersecurity policy Policymaking Steps

[3] underscore the importance of tailoring the cybersecurity approach to each organisation's unique cyber platform, necessitating the development of policies that guide users in utilizing technology. As articulated by the authors [3], the primary function of a cyber policy lies in securing business continuity by safeguarding critical areas. The following delineates the steps for formulating effective cybersecurity policies.

2.1 Cybersecurity policy Formulation step

Policy formulation is a process of identifying a problem and finding priorities and actions on a specific problem such as vulnerability and threats in the organisation [5]. In the formulation process, agendas for the policy are defined and set [6]. Policy formulation affects both implementation and outcomes because the success or failure of a policy depends mainly upon the policy formulation process [5]. Therefore, research must be collected to provide appropriate and accurate information for this process.

The initial step in formulating a cybersecurity policy is to establish a clear rationale and priorities the security focus within your organisation, as articulated by [4]. Furthermore, according to [7], successful policy formulation requires addressing key questions such as: How will the identified problem be resolved? What are the primary concerns and objectives? What options exist to achieve these objectives? What are the pros and cons of each option? Additionally, what externalities, whether favourable or unfavourable, are associated with each choice?

Organisations need to develop proactive cybersecurity policies to combat the increasing threat of cyberattacks. [4] draw attention to international policies that are common to all countries, which can assist policymakers in developing cybersecurity plans. The authors list fourteen shared characteristics that are essential to the development of strong cybersecurity policies. These characteristics include telecommunication, net- work security, cloud computing, e-commerce, e-business, identity theft, smart grid, and privacy. With these characteristics, organisations can rely on them to formulate the necessary policies for securing the internet.

Types of cybersecurity policies.

Cybersecurity policies constitute a multifaceted framework crucial for fortifying the security infrastructure of any organisation. [4] identified ten fundamental aspects, aligning with the conclusions drawn by [3] from [8]. These include information handling, access control, data retention, data protection, cloud computing, email, physical security, and network security.

Moreover, [3] introduced additional elements derived from the amalgamation of these literatures, underscoring the importance of incorporating policies pertaining to password management, company-owned device usage, WIFI usage, incident response, business continuity, disaster recovery, and backup strategies into the broader spectrum of information security policy documents. The ensuing detailed explanation delves into the formulation of policies designed to combat cybersecurity threats comprehensively. Below are the types of cybersecurity policies.

Privacy Policy and Data Protection: [4] and [3] underscore the crucial nature of privacy policies within organisations. [9] defines privacy policy as a legal document dictating how an organisation handles information related to its customers, clients, employees, and partners. It encompasses the collection, storage, use, sharing, protection of data, and outlines user rights concerning their information. Moreover, [10]. stress that a privacy policy discloses how an organisation collects, manages, and discloses customer data, including sensitive personal information. Legislation, such as the Data Protection Bill in Namibia (Draft Data Protection Bill, 2021), emphasises the need to regulate information processing to safeguard individuals' fundamental rights, especially the right to privacy.

Email Security Policy: [10] highlights the pivotal role of an effective email security policy in safeguarding against cyber threats and employee misuse. Emails serve as prime targets for phishing attacks, where malicious links and attachments can compromise sensitive information and introduce malware into an organisation's systems [11], Establishing rules and expectations for corporate email usage helps mitigate security risks by educating users on proper email system practices.

Network Security Policy: Network security, as defined by [11], involves safeguarding networks and data from breaches and intrusions through various measures like access control, antivirus software, firewalls, encryption, and more. [12] emphasises the need for a well-crafted network security policy to protect company assets while allowing.

efficient employee workflow. This policy outlines device connectivity, data transmission, and permissions, thereby enabling the blocking of malicious activities without altering hardware or software configurations [13]. Additionally, [14] emphasize the least privilege principle within network security policies, restricting network permissions to only essential users and applications.

Wi-Fi Usage Policy: [15] discovered that 50% of employees are indifferent to how they use the organisation's internet, partly due to managers being unclear about expectations. Consequently, to ensure effective network security, a robust Wi-Fi usage policy becomes essential. This policy outlines protocols for accessing and utilizing the organisation's wireless network, promoting secure connectivity, and thwarting unauthorised access. As highlighted by [16], such a policy serves as a preventive measure, mitigating the risk of employees causing unintentional or intentional harm to the company or its reputation.

Physical Security Policy: According to [17], cyber physical systems security incorporates security into a variety of interconnected computing systems and neighbouring system architectures. According to [4], cyber security policy includes physical equipment security in addition to the security of facilities, data, and information. Physical safety policies encompass measures to safeguard physical assets, facilities, and resources. This includes access control mechanisms, surveillance systems, and procedures to prevent unauthorised access or damage to physical infrastructure.

Password Management Policy: Exploits targeting weak passwords remain prevalent across various industries, as indicated by [18]. A password management policy establishes clear guidelines for the creation, storage, and regular updating of passwords, aiming to ensure robust authentication and prevent unauthorised access to systems and data. The effectiveness of a password policy lies in adhering to best practices, providing users with guidance on creating secure passwords [19]. The study conducted by [18] highlights the continued existence of possible weaknesses in the methods used to create passwords for online accounts. Because of these flaws in cyber hygiene, many systems are vulnerable to brute force attacks. [4] advise building a strong password policy, using two factor authentication, and taking precautions to prevent physical access to information technology infrastructure to defend against attacks that target specific information.

Response to Incidents, Business Continuity, and Disaster Recovery Policies: These policies are designed to establish protocols and strategies for effectively managing security incidents, maintaining operational continuity during disruptions, and ensuring prompt recovery from catastrophic events to minimise both downtime and data loss [20]. According to [21], the primary objective of a business continuity policy is to comprehensively document the requirements for sustaining organisational operations on regular business days and during emergencies. It is possible to set sensible expectations.

for disaster recovery (BC/DR) and business continuity procedures when there is a clear policy that is closely followed.

As highlighted by [22] is it importance of an Incident Response (IR) plan in guiding teams through the recovery processes following a cyberattack or breach. Equipping a company with complete information about response procedures to various cyber incidents, such as disclosure of confidential information, asset theft or damage, unauthorised use of services, malware in the system, unauthorised modifications and access to organisational hardware and software, disruption of the network, or failure of critical servers, proves extremely beneficial. Also, [23] emphasise the integration of disaster risk management and business continuity into organisational security policies and practices through disaster risk informed investments. According to [22], it is imperative for every organisation across industries to adopt three essential policies: Business Continuity, ensuring the restoration of business operations after a disaster or incident; Disaster Recovery, focusing on the IT aspect of restoring technology; and Incident Response, addressing cybersecurity incidents promptly. [22] stresses the harmonious coordination of all three policies to safeguard the organisation from potential disruptions.

Cyber awareness and education policy: Recognizing that computer security extends beyond securing systems and networks, it encompasses the individuals who utilise these systems and how their behaviors can influence cyber exploitation [24]. [25] underscores the significance of security awareness training as an integral component in safeguarding companies against evolving and detrimental cybersecurity threats. For this training to be truly effective, the formulation of a comprehensive security awareness and education policy is imperative. The Cybersecurity Awareness and Education Policy serve as a guiding document for all employees, urging them to undergo regular training sessions. This policy outlines the necessary protocols and expectations for fostering a cyber-aware workforce. It not only emphasises the identification of potential cyber threats but also addresses the behavioral aspects that can contribute to a secure computing environment. The policy aims to cultivate a proactive and vigilant organisational culture, where employees are informed, engaged, and equipped to play an active role in the collective cybersecurity efforts of the company.

By promoting ongoing education and awareness initiatives, the Cyber Awareness and Education Policy ensures that employees are abreast of the latest cybersecurity trends, threats, and best practices. Regular training sessions may cover topics such as recognizing phishing attempts, safeguarding sensitive information, adhering to password policies, and understanding the importance of secure communication practices. Additionally, the policy may outline consequences for non-compliance and underscore the shared responsibility of every employee in upholding the organisation's cyber resilience. The Cyber Awareness and Education Policy serve as a cornerstone in establishing a robust cybersecurity culture within the organisation, where knowledge, aware- ness, and proactive engagement collectively contribute to fortifying the human element against cyber threats.

On that note, a cyber-security policy framework and procedural compliance in public organisation was developed to suit the managers of cybersecurity activities in an organisation [3], The framework was developed with 7 themes. A table below will delve into the themes elucidated in the framework proposed by [3].

Table 1. Lubua's cyber-security policy framework [3]

Themes	
Data Security	To ensure that data is accounted for, secure and private.
Internet and Network Services Governance	To govern internet usage and services, to manage emails, social media usage, and connection of personal devices on the organisation network.
Uses of Company-Owned Devices	To govern company-owned devices and use acceptable use policy.
Physical security	To the measures and precautions put in place to safeguard people, assets, infrastructure, and information from physical threats, unauthorised access, theft, vandalism, or damage.
Incident Handling and Reporting	To respond to incidents, report security incidents, and scan for
Monitoring and Compliance	To monitor and control accounts and Software licensing, manage patches, and Forensic Auditing and risk management.
Policy Administration	To administrate policy, review, and training of system users.

The items listed in table 1, which is the framework proposed by [3], indicate the various kinds of policies that an organisation can create and include in a single document that serves as a cyber security policies document.

The formulation of cybersecurity policies represents a crucial and multi-faceted process essential for fortifying an organisation's security infrastructure. As emphasised by [5], the formulation step is foundational, influencing both the implementation and out- comes of policies. Research plays a pivotal role in providing the necessary information for this process, ensuring that policies are well-informed and effective. In essence, the comprehensive range of cybersecurity policies, from their formulation to their practical implementation, represents an intricate yet indispensable approach to mitigating cyber threats, securing organisational assets, and cultivating a resilient cybersecurity culture.

2.2 Cybersecurity policy Implementation step

The crucial stage of the policy-making process where the decisions made in the early stages are implemented is referred to as policy implementation. According to [26], it denotes the conversion of policy plans which are frequently conceived during formulation into concrete acts and practices. Expanding on this idea, [27] emphasise the dynamic aspect of policy implementation, describing it as an all-encompassing series of actions carried out by people or organisations with authority. These initiatives are deliberately created to actualise the goals and objectives specified in the policy statements.

This is a complex process that involves a range of players, tools, and tactics to guarantee that policies are carried out successfully. Careful planning, resource allocation, stakeholder coordination, and ongoing monitoring and evaluation to assess progress and make required modifications are all essential for an implementation to be carried out effectively. The successful execution of policies requires the collaboration and commitment of numerous stakeholders, such as officials from the public and private domains and non-governmental organisations. To guarantee coherence and alignment towards accomplishing the policy goals, it is imperative to establish unambiguous communication channels and clearly defined roles and responsibilities [6].

Additionally, there are several difficulties and complexities involved in implementing policies. Various factors, including insufficient funding, competing agendas among involved parties, formalities in the administrative process, and unanticipated outside events, can hinder or change the implementation process. To ensure that the desired policy outcomes continue to be attainable in the face of these challenges, flexibility and adaptability are essential qualities [28]. Furthermore, for policies to be implemented effectively, encouraging ownership and buy in from those involved is just as crucial as following directions. Developing capacity, fostering participation, and encouraging ownership among stakeholders can all help to improve commitment and sustainability during the execution phase. Furthermore, selecting the best controls, comprehending organisational requirements, disseminating and managing policies, providing awareness training, and keeping an eye on user behavior are just a few of the difficulties that come with implementing cybersecurity policies successfully [29].

2.3 Cybersecurity policy Modification step

Policies must be regularly reviewed and revised to remain relevant and effective in a world that is changing quickly, especially when it comes to cybersecurity. As suggested by [1] the model for cybersecurity policy demands should place a strong emphasis on an ongoing focus on the formulation of policies and iterative evaluations. According to [6], policy modification entails a systematic and intentional process of reviewing earlier policy choices and making necessary modifications in response to input from various stakeholders affected by these policies.

Policy performance evaluation and feedback mechanisms shape the dynamic process of policy evolution through modification. Experiences and results from the employees offer vital insights that guide the creation and modification of future policies. These assessments frequently highlight areas in which policies might be insufficient, have unforeseen consequences, or fall short of their goals. As a result, as [1] points out, policymakers must modify their original plans to better meet new demands and address.

evolving issues. The significance of feedback loops in the policymaking cycle, as underscored by [30] cannot be overstated. Stakeholder feedback and the continuous as- assessment of policy performance are pivotal in shaping the trajectory of subsequent policy iterations. This iterative process enables policymakers to adopt policies based on empirical evidence, real-world experiences, and evolving circumstances, ensuring a more responsive and effective approach to governance.

Moreover, changes can be made at any point during the process thanks to the adaptability of policy modification. This adaptability gives interested parties in an organisation those who gain from or are impacted by a policy the ability to push for changes that will improve or maintain the policy's efficacy over time. The iterative process of policy modification, fuelled by ongoing assessment, input from stakeholders, and practical observations, is essential to preserving policies that are flexible, efficient, and in line with changing demands and issues in cybersecurity and other domains. Policy modification, therefore, is an ongoing and dynamic process that integrates feedback, evaluation, and the dissemination of information. It enables policymakers to continually re- fine policies in response to changing conditions, new insights, and the evolving needs of the stakeholders involved. Through this iterative process, policies can better adapt to the complex and dynamic environments in which they operate, ultimately enhancing their effectiveness and relevance. Moreover, information communications and technology (ICT) keep revolving every minute. Therefore, the emergence of new technologies such as artificial intelligence, machine learning, and blockchain requires relevant methods of addressing security concerns and that is why cyber policies need to be comprehensive and updated regularly to cater to new criminal developments [3]. Figure 1 depict the policy making steps as discussed in this study [6].

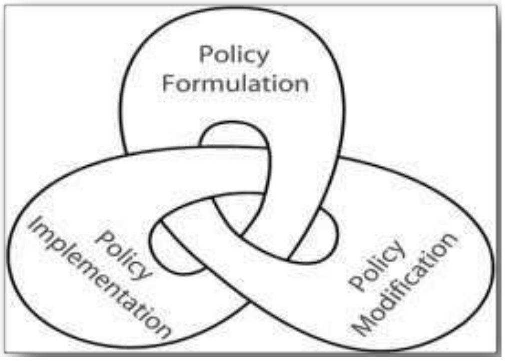


Fig. 1. Policymaking steps [6]

2.4 Policy Compliance step

Policy compliance is a cornerstone of effective cybersecurity within organisations. It involves the adherence to established cybersecurity policies and protocols by all stakeholders, both through technical and non-technical solutions. [31] encouraged the fostering a culture of security within an organisation is pivotal to ensuring policy compliance. In order to implement and maintain cybersecurity policies, technical solutions include the use of technological tools like intrusion detection systems, firewalls, multi- factor authentication, and encryption [32]. Non-technical solutions, on the other hand, involve aspects like creating awareness among employees, providing regular training sessions, and instituting clear guidelines and consequences for non-compliance.

Moreover, once a policy has been implemented, it is crucial to ensure that it is communicated effectively to all relevant parties and the importance of policy dissemination and comprehension among employees is ensured. This communication process in- volves making policies readily available and providing clear explanations to ensure that every individual comprehends their roles and responsibilities in adhering to and up- holding the policy. Policy compliance thus becomes essential, fostering a culture of adherence and accountability within the organisation.

A robust culture of policy compliance not only ensures that employees understand and adhere to cybersecurity policies but also serves as a bulwark against potential vulnerabilities and threats. It cultivates a shared responsibility toward safeguarding organisational assets and fosters a proactive approach to cybersecurity governance. Addition- ally, regular assessments, audits, and monitoring mechanisms are vital to evaluating and ensuring ongoing compliance with cybersecurity policies. In a similar vein, adhering to cyber security frameworks guarantees risk mitigation, threat prevention, and mechanism efficacy [1]. In the processes proposed by [6] the fourth process which is policy compliance has not been included.

However, the literature of [1][31] have identified the need for the policy formulation processes to include policy compliance as part of the cybersecurity processes. If the compliance process were to be integrated into the cybersecurity policy lifecycle process, the processes for cybersecurity policies would look like the one shown in figure 2 below:

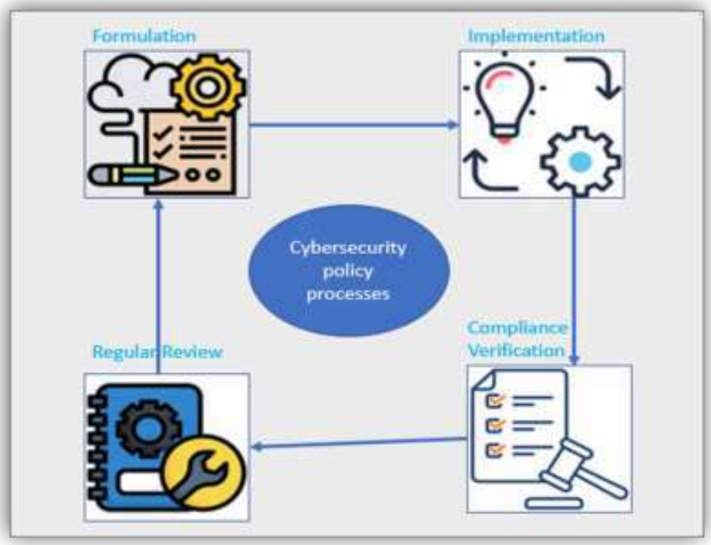


Fig. 2. The four processes of an effective Cybersecurity policy

The figure 2, shows that to fortify its digital defences, an organisation must first undertake the crucial task of developing a comprehensive cybersecurity policy. This foundational document serves as a strategic roadmap, outlining the principles and guidelines that will govern the organisation's approach to safeguarding its digital assets [4].

The formulation of the cybersecurity policy marks the initial step in the organisation's commitment to secure its information systems and sensitive data. Once this policy framework has been meticulously crafted, the next imperative is its seamless implementation throughout the organisation. However, it is paramount to ensure that, prior to implementation, the policy is thoroughly vetted for compliance with industry standards, legal regulations, and internal protocols. After the successful implementation of the cybersecurity policy, a dedicated team, typically led by the Chief Security Officer (CSO) or Chief Information Officer (CIO), assumes the responsibility of ensuring strict adherence to the established guidelines [31]. This team plays a pivotal role in educating and training all staff members and stakeholders on the intricacies of the policy. To track compliance, spot possible weaknesses, and quickly address any deviations, audits and assessments are carried out on a regular basis [27].

The dynamic nature of technology and the continuous evolution of cyber threats necessitate periodic reviews of the cybersecurity policy. Regular assessments provide opportunities to identify and address emerging risks, ensuring that the policy remains adaptive and resilient [29]. In this ever-changing landscape, modifications to the cyber- security policy become not just a possibility but a strategic imperative, allowing the organisation to stay ahead of potential threats and vulnerabilities. The repeated process of formulation, implementation, compliance verification, and regular review, coupled with the vigilant oversight of a dedicated team, establishes a robust cybersecurity framework. This framework not only safeguards the organisation's digital infrastructure but also positions it to respond effectively to the evolving challenges presented by the rapidly changing cyber landscape.

3 Conclusion

As organizations navigate the complex landscape of cybersecurity threats, the development and implementation of robust cybersecurity policies are paramount. This paper has clarified the essential steps involved in the policymaking process, underscoring the importance of tailoring policies to meet the specific needs and Challenges of organisations. By integrating policy compliance into the cybersecurity lifecycle, organizations can establish a proactive approach to cybersecurity governance, ensuring adherence to established guidelines and protocols. Through regular review and modification, organizations can adapt to evolving threats and technologies, thereby fortifying their digital defences, and maintaining resilience in the face of cyber adversaries.

References

1. Alqahtani, F. H. (2016). Developing an Information Security Policy: A Case Study Approach. Procedia Computer Science, 124, 691-697. https://doi.org/10.1016/j.procs.2017.12.206
2. Li, L., He, W., Xu, L., Ash, I., Anwar, M., & Yuan, X. (2019). Investigating the impact of cybersecurity policy awareness on employees' cybersecurity behavior. International Journal of Information Management, 45, 13-24. https://doi.org/10.1016/j.ijinfomgt.2018.10.017
3. Lubua, E. W. & Pretorius, P. D. (2019). Cyber-security Policy Framework and Procedural Compliance in Public Organisations. Proceedings of the International Conference on Industrial Engineering and Operations Management Pilsen, Czech Republic, July 23-26, 2019
4. Mishra, A., Alzoubi, Y. I., Anwar, M. J., & Gill, A. Q. (2022). Attributes impacting cybersecurity policy development: Evidence from seven nations. Computers & Security, 120, 102820. https://doi.org/10.1016/j.cose.2022.102820
5. Hansson-Forman, K., Reimerson, E., Bjärstig, T. & Sandström, C. (2021) A view through the lens of policy formulation: the struggle to formulate Swedish moose policy, Journal of Environmental Policy & Planning, 23:4, 528-542, DOI: 10.1080/1523908X.2021.1888700
6. Gender-sensitive Policies and Programmes (n.d). Unit 13 Policy Formulation and Development
7. Taylor, L. (2001). Seven elements of highly effective security policies. Retrieved from https://www.zdnet.com/article/seven-elements-of-highly-effective-security-policies/
8. Wright, Gavin. (2023, November). Definition privacy policy. TechTarget. https://www.techtarget.com/whatis/definition/privacy-policy#:~:text=A%20privacy%20policy%20is%20a,in%20connection%20to%20the%20data.
9. Chua, H. N., Herbland, A., Wong, S. F., & Chang, Y. (2017). Compliance to personal data protection principles: A study of how organizations frame privacy policy notices. Telematics and Informatics, 34(4), 157-170. https://doi.org/10.1016/j.tele.2017.01.008

10. Froehlich, Andrew. (2023, 30 October). What an email security policy is and how to build one. TechTarget. https://www.techtarget.com/searchsecurity/tip/Why-you-need-an-email-security-policy-and-how-to-build-one
11. Checkpoint (n.d). Security Policy. Retrieved from https://sc1.checkpoint.com/documents/R81/WebAdminGuides/EN/CP_R81_NextGenSecurityGateway_Guide/Topics-FWG/Security-Policy.htm#:
12. EC-Council (3 June 2022). Understanding and Designing Strong Network Security Policies. Retrieved from https://www.eccouncil.org/cybersecurity-exchange/network-security/understand-design-implement-network-security-policies/
13. Barrera, D., Molloy, I. & Huang, H. (2018). Standardizing IoT Network Security Policy Enforcement. Workshop on Decentralized IoT Security and Standards (DISS) 2018., San Diego, CA, USA. ISBN 1-891562-51-7. https://dx.doi.org/10.14722/diss.2018.23007
14. Katsis, C., Cicala, F., Thomsen, D., Ringo, N., & Bertino, E. (2021, June). Can I reach you? do I need to? new semantics in security policy specification and testing. In Proceedings of the 26th ACM Symposium on Access Control Models and Technologies (pp. 165-174).
15. Bhuwal, H. (2020, June 29). Efficient internet usage policy to set employees up for success. https://empmonitor.com/blog/internet-usage-policy/
16. ComTech Computer Services, Inc. (2015, August 14). Why your company needs an Internet. Use Policy. https://www.comtech-networking.com/blog/item/152-why-your-company-needs-an-internet-use-policy/
17. DiMase, D., Collier, Z.A., Heffner, K. et al. (2015). Systems engineering framework for cyber physical security and resilience. Environ Syst Decis 35, 291–300 (2015). https://doi.org/10.1007/s10669-015-9540-y
18. Hall, R. C., Hoppa, M. A., & Hu, Y. H. (2023, March). An Empirical Study of Password Policy Compliance. In Journal of The Colloquium for Information Systems Security Education (Vol. 10, No. 1, pp. 8-8).
19. Lee, K., Sjöberg, S., & Narayanan, A. (2022). Password policies of most top websites fail to follow best practices. In Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022) (pp. 561-580).
20. Niemimaa, M., Järveläinen, J., Heikkilä, M., & Heikkilä, J. (2019). Business continuity of business models: Evaluating the resilience of business models for contingencies. International Journal of Information Management, 49, 208-216. https://doi.org/10.1016/j.ijinfo- mgt.2019.04.010
21. Sullivan, E. (May 2022). business continuity policy. https:// asterrecovery/definition/business-continuity-policy#:~:text=The%20goal%20of%20a%20business,(BC%2FDR)%20processes.
22. Convocar, J. K. (2022, July 28). Disaster Recovery vs. Business Continuity vs. Incident Response Plans. https://www.itsasap.com/blog/disaster-recovery-vs-business-continuity-vs-incident-response
23. Kato, M., & Charoenrat, T. (2018). Business continuity management of small and medium sised enterprises: Evidence from Thailand. International Journal of Disaster Risk Reduction, 27, 577-587. https://doi.org/10.1016/j.ijdrr.2017.10.002
24. Zwilling, M., Klien, G., Lesjak, D., Wiechetek, Ł., Cetin, F., & Basim, H. N. (2022). Cyber security awareness, knowledge and behavior: A comparative study. Journal of Computer Information Systems, 62(1), 82-97. DOI: 10.1080/08874417.2020.1712269
26. Ahola, M. (2023). How to make a good security awareness training policy? (with free template). https://blog.usecure.io/what-makes-a-good-security-awareness-training-policy

27. Cerna, Lucia. (2013). The Nature of Policy Change and Implementation: A Review of Different Theoretical Approaches. Organisation for economic co-operation and development
28. Bullock, H.L., Lavis, J.N. (2019). Understanding the supports needed for policy implementation: a comparative analysis of the placement of intermediaries across three mental health systems. Health Res Policy Sys 17, 82. https://doi.org/10.1186/s12961-019-0479-1
29. Flowerday, S. V., & Tuyikeze, T. (2016). Information security policy development and implementation: What, how and who. Computers & Security, 61, 169-183. https://doi.org/10.1016/j.cose.2016.06.002
30. Mishra, A., Alzoubi, Y. I., Gill, A. Q., & Anwar, M. J. (2021). Cybersecurity Enterprises Policies: A Comparative Study. Sensors (Basel, Switzerland), 22(2). https://doi.org/10.3390/s22020538
31. Guws Medical, (2021, October). Distinguishing Policy Modification from Policy Initiation. guwsmedical.com. https://www.guwsmedical.info/health-policy/distinguishing-policy-modification-from-policy-initiation.html
32. Yusif, S., & Hafeez-Baig, A. (2021). A conceptual model for cybersecurity governance. Journal of applied security research, 16(4), 490-513. DOI: 10.1080/19361610.2021.1918995
33. Mullet, V., Sondi, P., & Ramat, E. (2021). A review of cybersecurity guidelines for manufacturing factories in industry 4.0. IEEE Access, 9, 23235-23263.

Behavioural Predictors that Influence Digital Legacy Management Intentions among Individuals in South Africa

Jordan Young[1[0000-0000-0000-0000]], Ayanda Pekane[1[0000-0002-4756-6357]] and Popyeni Kautondokwa[1[0000-0001-9001-7313]]

[1] University of Cape Town, Cape Town, 7700, South Africa

Abstract. An emerging phenomenon, digital legacy management explores the management of digital data individuals accumulate throughout their lifetime. With the integration of digital systems and data into people's daily lives, it becomes crucial to understand the intricacies of managing data to eventually become one's digital legacy. This can be understood by investigating the significance of behavioural predictors in shaping digital legacy management. The objective of this study is to explore how behavioural predictors influence the intentions of individuals in South Africa towards managing their digital legacy. This entailed: 1) investigating the impact of attitude, subjective norms, and perceived behavioural control on these intentions; 2) exploring the perceived usefulness of digital legacy management systems; and lastly 3) understanding the implications of response cost and task-technology fit on individuals' inclinations towards digital legacy planning. Data were collected (n = 203 valid responses) from South African residents using an online survey and analysed using partial least squares structural equation analysis (PLS-SEM). Results indicate that attitudes, peer opinions, personal resources and skills are significant positive influences on digital legacy management intention. Recognizing and understanding these behavioural predictors is key when developing region-specific and culturally sensitive digital legacy management tools, awareness campaigns and policies. Furthermore, it could pave the way for more tailored strategies, ensuring effective transfer of post-mortem data, reducing potential conflicts, and providing clarity when dealing with post-mortem data.

Keywords: Digital Legacy, Digital Legacy Management, Personal Data Management Predictors, Digital Afterlife, Post-Mortem Data.

1 Introduction

During the 21st century, the value of digital objects has drastically increased becoming what we now refer to as digital assets [13] . Historically, legacies typically comprised of tangible items that were passed down from one individual to another, yet these physical items have been largely replaced by digital equivalents [47]. Digital data often serves as an extension of oneself and can be of high importance to individuals [12]. As people spend more time in their virtual lives, it is important to protect what we leave behind [7]. However, the shift towards digital versions of physical items has not been followed by a corresponding shift in the way these digital items are managed and passed down to future generations [47].

Digital legacy refers to the accumulated digital information, across various platforms and formats, left behind by an individual after their death, which includes content they created, shared, or that pertains to them [12, 13, 47]. Very few people have planned for their digital legacy as it is a relatively new concept [18]. The concept of digital legacy has recently been brought to light through social media and digital platforms such as Facebook [13] where guidelines for deceased user accounts are still emerging [9]. Managing one's digital legacy and understanding these behavioural predictors impacts: 1) the bereaved who may seek access or closure; 2) the way digital platforms handle one's data after one's death; and 3) the deceased's privacy, as their personal content could be accessed in undesired ways.

Due to the increase in data stored online and how frequently users interact online, there is growing concern about how valuable digital assets can be managed and passed on to loved ones [31]. Even death-aware people have not considered their own digital legacy [44]. The consequences of not managing one's digital legacy could include copyright infringement, invasion of privacy, theft of identity, and loss of private data [29]. Recognizing behavioural predictors is key to understanding how individuals approach digital legacies and tailored digital legacy management strategies can be developed, reducing conflicts and uncertainties. The implications of unmanaged digital legacies range from emotional distress of the bereaved to potential legal issues.

The objective of this study is to understand the behavioural predictors that influence digital legacy management intentions among individuals in South Africa. Understanding behavioural predictors is vital for the future of effective digital legacy management. As South Africa's digital landscape rapidly expands, it becomes imperative to address the complexities of managing and safeguarding digital assets for future generations [3, 6].

2 Literature Review

2.1 Digital Legacy Platforms

In recent years there has been a rise in digital afterlife research and many digital legacy platforms have been developed such as Afternote [30, 32]. Afternote is a digital platform that enables users to save their personal history, leave messages for loved ones, and record their final wishes [29]. These digital legacy platforms, designed for death-related practices that aim to preserve the memories of loved ones currently have small user bases; however, mainstream social media platforms have been gaining popularity amongst users for grieving and mourning-related posts [12]. Many startups have been capitalizing and thriving in the end-of-life sector, which is expected to continue as technology infiltrates society [29]. The digital afterlife industry ranges from small business applications like Afternote to larger ones like Facebook. Some digital afterlife platforms are free, and others require a fee, but these fees are based on a rate not intended for the standard of living in the Global South [29].

It is important not to blindly adopt unsuitable solutions from Western society, which refers to North American and European regions with predominantly Eurocentric values and perspectives. Despite this, popular platforms in the Global South, specifically in the context of South Africa are currently based on Western-influenced policies and guidelines.

2.2 Social Networking Sites

With the increase in Social Networking Sites (SNS) and trends suggesting that the rise in technology usage will continue, it is crucial to understand the long-term consequences for users in the context of end-of-life [8]. Some view the interaction and information on SNS as one's 'digital soul' that will become one's digital legacy [8]. Previous research has been done on the social relationships that form between the bereaved and deceased through social media [34]and how SNS manage their responsibility as a digital memorial site [33]. These SNS create a space for the bereaved to receive social support from strangers, fellow sufferers and other loved ones at any time and place rather than having to rely on someone in close physical proximity [5].

2.3 Managing Digital Assets

Users commonly struggle to manage their data due to: 1) large amounts of data they possess; 2) lack of motivation; and 3) the time and effort needed [12]. Despite the challenges, individuals find value in their digital assets, offering a sense of pride and fulfilment, making the management of their digital legacy worthwhile due to the significant value inherited digital assets can hold [12, 47]. There is a general societal concern for the management of digital legacies [10]. In a study done by [26] a little more than half (56%) of the respondents were worried about the management of their personal technologies after their death.

2.4 Perceptions

Awareness about Digital Legacy.

There is a lack of awareness regarding the management of digital assets and the potential to leave behind resource-intensive digital waste after one's death [29, 31]. [31] found that many online users have not considered their digital legacy but believed it should be considered in the future. Even among people who have a legal will, 70% had no clear understanding of what would happen to their digital assets after their passing [13]. [8] found that many students had never considered their own digital legacy unless they had experienced the passing of loved ones with active SNS.

Attitudes towards Digital Legacy.

Western society has traditionally viewed death as taboo. Although recent developments in technology have shifted people's attitudes towards death, individuals still tend to avoid making end-of-life decisions [28]. This can be explained by terror management theory that suggests that individuals avoid these decisions due to their belief that they are not going to die soon. This may explain the low usage of digital asset management tools. Additionally, people are generally not enthusiastic about planning for their death as it can involve tedious planning and thoughtful evaluation of what to leave behind [12]. People are motivated to plan their digital legacy if they view digital data as a gift as opposed to a burden, as it can then be framed as a meaningful process [12].

To safeguard one's digital assets, it is essential that people are educated about proper planning [8].

Preferences towards Digital Legacy.

There are varying preferences regarding digital legacy and how digital remains are managed after one's death. [25] researched how young people, who represent the internet generation, comprehend death and how that shapes their digital posthumous interaction. They found that 53.8% of the participants wished to leave a posthumous message that will be displayed after their death [25]. Similarly, the respondents in [44] study expressed a desire towards ensuring their valued digital artefacts are preserved, not only for themselves but for their loved ones. In contrast, [8] found that the majority of participants desired that their online digital remains are deleted, with only 24.4% wanting their own profiles to be active after their passing. Although the majority of these remains are text-based content and photographs at present, over time it is likely to expand to various other types of content as more applications and services are moved to the cloud.

The control and management of digital assets after death is another area of concern for individuals. Some people prefer to control which of their digital assets are to be kept or deleted, while others prefer automated alternatives [43]. A study found that 45%-50% of individuals preferred that someone is granted access to their personal email, social platforms, and digital accounts after their passing [18]. 31%-36% wished that all access be denied to their digital assets, and the remaining individuals preferred partial access [18].

Religious beliefs.

Cultural and religious beliefs influence how people view [25]. This can provide a guideline for SNS and digital legacy management platforms to respect users' digital legacy preferences and beliefs. Some research has been done on how rituals and practices can be enhanced through blending physical and digital interaction [34]. Religious farewell rituals demonstrate individuals' desire to maintain a connection with deceased loved ones [25]. Death is a social and cultural construct with specific sets of values and meanings. These cultural beliefs, rooted in religion, determine appropriate interactions with the dead [25]. [25] found two opposing ideas about death. Some believe that there is an abstract life beyond death while others view death as the end.

2.5 Challenges in Post-Mortem Data

Inheriting digital assets is complex due to the lack of established social, cultural, and religious guidelines [13, 34]. There are four main challenges associated with digital legacies after someone dies: 1) Problems with the transfer and access to assets due to authentication; 2) email access and password protection; 3) how to ensure the longevity of digital assets, and 4) concerns with the level of understanding of digital legacy terminology [13]. There is a trade-off between a service provider having access to one's data while providing privacy in exchange; however, once a user passes, the provider continues to have access [18]. This raises questions about post-mortem privacy rights, which is a person's right to control their digital legacy and assets after death [18].

Policies And Regulations Issues.

In Common Law jurisdictions, privacy rights end when a person dies, but in the digital age online service providers continue to store and control the data, highlighting one of the challenges with post-mortem data [28]. When someone dies, friends and family need the permission of service providers to gain access to data. This has resulted in legal complications and the few cases that have gone to court resulted in mixed outcomes [28]. Service providers often indicate in their terms and conditions that they have no legal obligation to grant access to data when a user passes away [47]. This is because their revenue comes from active users, which makes the allocation of resources to manage inactive accounts redundant [47]. The shift towards digital transformation, which includes online distribution of personal and sensitive information might cause stress for those receiving a digital legacy and for the curators thereof [33]. This is due to the fact that existing practices have not been adapted to the digital domain, even though its significance continues to grow [33]. The deceased's privacy rights can raise legal concerns regarding the extent of access an heir can have to an account [47]. Legally, three parties are affected by the contents of digital legacies: 1) the deceased individual who agreed to the terms of service; 2) the services they signed up for that established the rules; and 3) the heirs who may seek access to the deceased's account [47]. If the digital executor is left with account details, they have the potential to act as the deceased user and invade their privacy. This raises the ethical question of whether heirs should or should not be given access to deceased accounts.

3 Conceptual Background

Fig.1 outlines the adapted conceptual model which includes constructs from protection motivation theory, technology acceptance model, task-technology fit and theory of planned behaviour [2, 23, 37]. PMT explores the motives for one’s digital legacy protective behaviours. TAM addresses the acceptance of new technology, while the TTF model focuses on the fit between task characteristics and technology characteristics. Finally, TPB integrates attitudes, societal norms, and perceived controls to influence one's behavioural intentions.

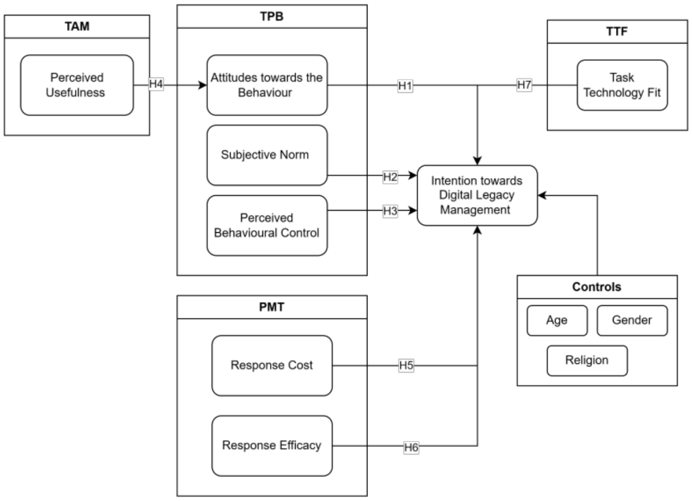


Fig. 1. Conceptual model adapted from [2, 23, 37].

3.1 Hypotheses

Hypotheses are important in deducing from theory and in subsequently, testing the hypotheses to come to a conclusion. In Table 1, the hypotheses developed for the study are presented.

Table 1. Hypotheses developed for the study.

Hypotheses	Supporting Literature
1. Attitudes towards digital legacy management outcomes have a positive influence on individuals' intentions to manage their digital legacy.	Attitudes and intentions can be influenced by ideas about the outcomes of a conduct .[2]
2. Subjective norms of friends and family have a positive influence on digital legacy management intentions.	Subjective norm is a normative cognition that represents a person's assessment of whether significant people want them to engage in the target activity as well as their drive to comply with these others [15]. The stronger the positive perception of the subjective norm with respect to a behaviour, the more likely it is that there will be an intent to perform the behaviour [2].
3. Perceived behavioural control has a positive influence on digital legacy management intentions.	The combination of intentions and perceptions of behavioural control significantly contributes to the variation observed in behavioural intention [2]. Perceived behavioural control represents the capabilities and resources users possess, and lacking these essentials, their

Hypotheses	Supporting Literature
	intention towards a behaviour will be diminished [24].
4. Perceived usefulness of digital legacy management systems has a positive influence on attitudes towards digital legacy management intentions.	According to the Technology Acceptance Model (TAM), an individual's intention to accept technology is directly influenced by attitude which is influenced by perceived usefulness [1].
5. Response cost has a negative influence on digital legacy management intentions.	A higher perceived response cost will result in lower likelihood of individuals engaging in a specific protective behaviour [42].
6. Response efficacy has a positive influence on digital legacy management intentions.	Individuals who perceive a measure as effective are more likely to develop an intention to adopt it [19].
7. Task-technology fit has a positive influence on intention towards digital legacy management.	Studies suggest that task-technology fit leads to effective utilization. However, such utilization and enhanced performance cannot be achieved until the intention is realized [11].

4 Methodology

The study used a positive paradigm to employ scientific methods and adopted a deductive approach, employing Theory of Planned Behaviour, Technology Acceptance Model, Protection Motivation Theory and Task Technology Fit to formulate testable hypotheses. Utilizing a descriptive research design, the study examined the concepts and relationships related to digital legacy, a relatively new field, in order to establish a foundational understanding that can inform future studies.

4.1 Data Collection

A survey questionnaire was used to gather data from participants. This research instrument provided an adequate method of collecting data from a large sample while eliminating the need for the researcher to be present. This questionnaire was distributed to South African residents over the age of 18 using the platform Qualtrics. The questionnaire consisted of closed-ended structured questions that have been developed from previous studies, for simplicity, all questions were transformed to 5-item Likert with item responses ranging from 1 (strongly disagree) to 5 (strongly agree).

5 Data Analysis and Findings

The survey questionnaire was created, disseminated, and recorded using Qualtrics over a period of two weeks. Out of the initial 228 responses, 25 incomplete responses were removed, leaving a final sample of 203 valid responses. Data was analyses using PLS-SEM in SmartPLS.

5.1 Demographics Description

The demographics for age were segmented into five groups: 18-30, 31-40, 41-50, 51-60, and 61+. The majority of respondents (33.0%) were between the ages of 18-30 years old. 13.8% of respondents fell within the 31-40 age bracket. 12.3% of the participants were aged between 41-50 and 16.3% of respondents were between 51-60 years old. Lastly, 24.6% of respondents were 61 years or older.

5.2 Internal Consistency Reliability

Composite reliability was used to measure the internal consistency reliability, which is the preferred measure in PLS-SEM as it does not assume that each indicator is equally reliable [16, 46]. Composite reliability of values above 0.70 are satisfactory, as seen in Table 2, all constructs pass the composite reliability check.

Table 2. Composite Reliability.

Construct	Composite Reliability
Attitudes	0.920
Intentions	0.905
Perceived Behavioural Control	0.893
Perceived Usefulness	0.862
Response Cost	0.796
Response Efficacy	0.831
Subjective Norms	0.873
Task Technology Fit	0.933

5.3 Convergent Validity

Convergent Validity is tested using Average Variance Extracted (AVE) which indicates whether the latent construct can explain more than half of its indicators' variances [17]. An AVE of 0.50 or more indicates a sufficient degree of convergent reliability and will be accepted [16]. As seen in Table 3 the AVE extracted from the model shows that all constructs pass [36].

Table 3. Average Variance Extracted.

Construct	Average Variance Extracted (AVE)
Attitudes	0.743
Intentions	0.760
Perceived Behavioural Control	0.807
Perceived Usefulness	0.678
Response Cost	0.567
Response Efficacy	0.622
Subjective Norms	0.696
Task Technology Fit	0.823

5.4 Hypotheses Test Results

To assess the validity of the hypothesized relationships in the model, the path coefficients, t-values, and p-values are examined [4, 36]. Path coefficients, which typically fall between -1 and +1, signify the strength and direction of the relationship between constructs. Values closer to +1 suggest strong positive relationships, whereas values closer to -1 indicate strong negative relationships [36]. The p-value indicates the likelihood that a statistical outcome would occur due to chance [4]. A p-value lower than 0.05 is considered statistically significant at a 5% level, confirming the hypothesized relationship. T-values greater than 1.96 in two-tailed testing indicate a 5% level of statistical significance [16]. Table 4 shows each hypothesis path coefficients and its significance.

Table 4. Path coefficients of the structural model and significance testing results.

Hypothesis	**Path**	**Path Coefficient**	**t-value**	**p-value**	**Supported**
H1	ATT -> INT	0.227	3.323	0.001	**Yes**
H2	SN -> INT	0.285	3.850	0.000	**Yes**
H3	PBC -> INT	0.345	5.658	0.000	**Yes**
H4	PU -> ATT	0.352	4.919	0.000	**Yes**
H5	RC -> INT	-0.140	2.243	0.025	**Yes**
H6	RE -> INT	-0.006	0.095	0.924	No
H7	TTF -> INT	0.045	0.611	0.541	No

6 Discussion

Based on the results, it was found that the attitude towards digital legacy management has a significant causal relationship on intention. This corroborates with the TPB framework by [24] that looks at users' behavioural intentions towards a digital communication tool. Furthermore, the results demonstrate that the perceived usefulness of managing one's digital legacy positively influences the attitude towards digital legacy management. This is in accordance with studies that explore individuals' acceptance of the usage of new technologies [1].

The findings indicate that subjective norms have a significant causal relationship with the intention to manage one's digital legacy. The results are similar with studies from [35, 38] that explore the digital management of banking. It highlights the importance of societal and peer perspectives in shaping an individual's intention towards digital legacy management. The results show that the demographic control gender has a significant influence on subjective norms. This could imply that societal norms affect genders differently with regards to digital legacy management. This is in line with gender studies that found peer influence to have a greater influence on women [27, 45].

The perceived behavioural control was also found to have a significant causal relationship on intention. This finding is consistent with a study done by [24] looking at users' behavioural intentions towards a digital communication tool. It implies that an individual's confidence in their capabilities and the resources play a role in influencing their intentions towards digital legacy management.

The results show that the response cost negatively influences digital legacy management intentions, which corroborates with previous literature looking at information security behaviours [21, 40, 42]. When individuals perceive a high cost associated with managing their digital legacy, they are less inclined to engage in digital legacy management. However, the results demonstrate that response efficacy's influence on intentions was insignificant and in the opposite direction than the hypothesised positive direction. This is contrary to numerous studies findings [14, 19, 40], although, there are instances where the anticipated positive influence of response efficacy on intentions were not supported [39, 41]. For individuals to actively manage and protect their digital assets and online presence after death, it is expected that they recognize the advantages of safeguarding these digital legacies [42]. In the demographic analysis, Figure 4 reveals that 82% of participants are either unfamiliar with digital legacy or are indifferent towards digital legacy management. This could be why response efficacy was not supported.

The findings revealed that the task-technology fit is insignificant and does not support intentions towards digital legacy management. This is contrary to previous studies looking at the perceived fit between technology and a task and the user's intention to use that technology [11, 20, 22]. This suggests that there is a poor alignment between individual needs in this domain and the technology solutions available. Another problem similar to response efficacy, is that participants are unfamiliar with the technologies designed for digital legacy management, preventing them from effectively understanding task-technology fit.

7 Conclusion

The rapid growth of digitalization in the past decades has highlighted the significance of managing one's digital legacy. This pertains to managing one's digital assets, including social media internet interactions and personal digital data. Understanding the intricacies of digital legacy management has become crucial for individuals, loved ones of the deceased and digital service providers. With South Africa's increasing digital integration into one's daily life, this study provides a unique context for understanding digital legacy management due to this country's digital divide and distinct digital legacy policies.

This study contributes to literature using the Theory of Planned Behaviour, Technology Acceptance Model, Protection Motivation Theory and Task-Technology Fit. By employing a multi-theoretical approach, this study has provided a broad perspective on what behavioural predictors influence individuals' intention to manage their digital legacy and how they do so in South Africa. The behavioural predictors are Attitudes, Subjective Norms, Perceived Behavioural Control, Perceived Usefulness and Response

Cost, which have all shown to be significant for digital legacy management intentions. These valuable insights hold significant potential for influencing platform design, awareness campaigns, and shaping policy frameworks tailored to a South African context. This research serves as a foundation for academicians, digital platforms, and technology companies in the digital legacy management field. Lastly, this paper emphasizes the need for platform developers to reassess and standardize post-mortem data designs and policies according to user preferences, resources and local regulations.

References

1. Abu-Dalbouh HM (2013) A questionnaire approach based on the technology acceptance model for mobile tracking on patient progress applications. Journal of Computer Science 9. doi: 10.3844/jcssp.2013.763.770
2. Ajzen I (1985) From intentions to actions: A theory of planned behavior. In: Action Control: From Cognition to Behavior. pp 11–39
3. Aruleba K, Jere N (2022) Exploring digital transforming challenges in rural areas of South Africa through a systematic review of empirical studies. In: Scientific African
4. Bhattacherjee A (2012) Social Science Research: Principles, Methods, and Practic-es
5. Blaß M, Graf-Drasch V, Schick D (2022) Grief in the Digital Age-Review, Synthesis, and Directions for Future Research. Wirtschaftsinformatik
6. Bosch T (2010) Digital journalism and africa online public spheres in south. Communication 36. doi: 10.1080/02500167.2010.485374
7. Braman J, Dudley A, Vincenti G (2011) Death, social networks and virtual worlds: A look into the digital afterlife. In: Proceedings - 2011 9th International Conference on Software Engineering Research, Management and Applications, SERA. pp 186–192
8. Braman J, Vincenti G, Dudley A, Wang Y, Rodgers K, Thomas U (2013) Teaching about the Impacts of Social Networks: An End-of-Life Perspective Nordiana-Shah publication. 240–249
9. Brubaker JR, Hayes GR, Dourish P (2013) Beyond the grave: Facebook as a site for the expansion of death and mourning. The Information Society 29:152–163. doi: 10.1080/01972243.2013.777300
10. Cerrillo-i-Martínez A (2018) How do we provide the digital footprint with eternal rest? Some criteria for legislation regulating digital wills. Computer Law & Security Review 34:1119–1130. doi: 10.1016/j.clsr.2018.04.008
11. Chen, Huang (2017) The effect of task-technology fit on purchase intention: The moderating role of perceived risks. Journal of Risk Research 20. doi: 10.1080/13669877.2016.1165281
12. Chen JX, Vitale F, McGrenere J (2021) What Happens After Death? Using a Design Workbook to Understand User Expectations for Preparing their Data. In: Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems. pp 1–13
13. Cook DM, Dissanayake DN, Kaur K (2019) The usability factors of lost digital legacy data from regulatory misconduct: older values and the issue of ownership. In: 2019 7th International Conference on Information and Communication Technology (ICoICT. pp 1–6
14. Doane AN, Boothe LG, Pearson MR, Kelley ML (2016) Risky electronic communication behaviors and cyberbullying victimization: An application of Protection Motivation Theory. Computers in Human Behavior 60. doi: 10.1016/j.chb.2016.02.010
15. Hagger MS, Chatzisarantis NLD, Biddle SJH (2002) A Meta-Analytic Review of the Theories of Reasoned Action and Planned Behavior in Physical Activity: Predictive Validity and the Contribution of Additional Variables. Journal of Sport and Exercise Psychology 24:3–32. doi: 10.1123/jsep.24.1.3

16. Hair J, Ringle C, Sarstedt M (2011) PLS-SEM: Indeed a Silver Bullet. Journal of Marketing Theory and Practice 19:139–152. doi: 10.2753/MTP1069-6679190202
17. Hair J, Sarstedt M, Ringle C, Gudergan S (2021) Advanced issues in partial least squares structural equation modeling (PLS-SEM. 4(1). Sage publications
18. Holt J, Nicholson J, Smeddinck JD (2021) From personal data to digital legacy: Exploring conflicts in the sharing, security and privacy of post-mortem data. In: The Web Conference 2021 - Proceedings of the World Wide Web Conference, WWW 2021
19. Ifinedo P (2012) Understanding information systems security policy compliance: An integration of the theory of planned behavior and the protection motivation theory. Computers and Security 31. doi: 10.1016/j.cose.2011.10.007
20. Li Y, Yang S, Zhang S, Zhang W (2019) Mobile social media use intention in emergencies among Gen Y in China: An integrative framework of gratifications, task-technology fit, and media dependency. Telematics and Informatics 42. doi: 10.1016/j.tele.2019.101244
21. Liang H, Xue Y (2010) Understanding security behaviors in personal computer usage: A threat avoidance perspective. Journal of the Association for Information Systems 11. doi: 10.17705/1jais.00232
22. Lin WS (2012) Perceived fit and satisfaction on web learning performance: IS continuance intention and task-technology fit perspectives. International Journal of Human Computer Studies 70. doi: 10.1016/j.ijhcs.2012.01.006
23. Lu H, Yang Y (2010) Modelling the factors that affect individuals- utilisation of online learning systems: An empirical study combining the task technology fit model with the theory of planned behaviour. British Journal of Educational Technology 41. doi: 10.1111/j.1467-8535.2010.01054.x
24. Lu Y, Zhou T, Wang B (2009) Exploring Chinese users' acceptance of instant messaging using the theory of planned behavior, the technology acceptance model, and the flow theory. Computers in Human Behavior 25. doi: 10.1016/j.chb.2008.06.002
25. Maciel C, Pereira VC (2013) Social Network Users' Religiosity and the Design of Postmortem Aspects. In: 14th International Conference on Human-Computer Interaction (INTERACT. pp 640–657
26. Massimi M, Baecker RM (2010) A death in the family. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. pp 1821–1830
27. Miller JB (1986) Toward a New Psychology of Women, Beacon
28. Morse T, Birnhack M (2022) The posthumous privacy paradox: Privacy preferences and behavior regarding digital remains. New Media & Society 24:1343–1362. doi: 10.1177/1461444820974955
29. Mostafa M, Hussain F (2021) Transcending Old Boundaries: Digital Afterlife in the Age of COVID-19
30. Öhman C, Floridi L (2017) The Political Economy of Death in the Age of Information: A Critical Approach to the Digital Afterlife Industry. Minds and Machines 27:639–662. doi: 10.1007/s11023-017-9445-2
31. Peoples C, Hetherington M (2015) The cloud afterlife: Managing your digital legacy. In: 2015 IEEE International Symposium on Technology and Society (ISTAS. pp 1–7
32. Pereira HS, F. T, F. P, C., O. P, R. (2019). Exploring Young Adults' Understanding and Experience with a Digital Legacy Management System Journal of Interactive Systems 10. doi: 10.5753/jis.2019.553
33. Pfister J (2017) This will cause a lot of work.' - Coping with Transferring Files and Passwords as Part of a Personal Digital Legacy. In: Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing. pp 1123–1138

34. Pitsillides S (2019) Digital legacy: Designing with things. Death Studies 43:426–434. doi: 10.1080/07481187.2018.1541939
35. Puschel J, Mazzon JA, Hernandez JMC (2010) Mobile banking: proposition of an integrated adoption intention framework'. International Journal of Bank Marketing 28:389–409
36. Ringle C, Sarstedt M, Straub D (2012) Editor's Comments: A Critical Look at the Use of PLS-SEM. MIS Quarterly 36
37. Rogers RW (1975) A Protection Motivation Theory of Fear Appeals and Attitude Change. The Journal of Psychology 91. doi: 10.1080/00223980.1975.9915803
38. Sripalawat J, Thongmak M, Ngramyarn A (2011) M-banking in metropolitan Bangkok and a comparison with other countries'. Journal of Computer Information Systems 51:67–76
39. Thompson N, McGill TJ, Wang X (2017) Security begins at home": Determinants of home computer and mobile device security behavior. Computers and Security 70. doi: 10.1016/j.cose.2017.07.003
40. Tsai HYS, Jiang M, Alhabash S, Larose R, Rifon NJ, Cotten SR (2016) Understanding online safety behaviors: A protection motivation theory perspective. Computers and Security 59. doi: 10.1016/j.cose.2016.02.009
41. Vance A, Siponen M, Pahnila S (2012) Motivating IS security compliance: Insights from Habit and Protection Motivation Theory. Information and Management 49 3–4. doi: 10.1016/j.im.2012.04.002
42. Verkijika SF (2018) Understanding smartphone security behaviors: An extension of the protection motivation theory with anticipated regret. Computers and Security 77. doi: 10.1016/j.cose.2018.03.008
43. Vitale F, Odom W, McGrenere J (2019) Keeping and discarding personal data: exploring a design space. In: Proceedings of the 2019 on Designing Interactive Systems Conference. pp 1463–1477
44. Waagstein A (2014) An exploratory study of digital legacy among death aware people. Thanatos 3:46–67
45. White Baker E, Al-Gahtani SS, Hubona GS (2007) The effects of gender and age on new technology implementation in a developing country: Testing the theory of planned behavior (TPB. Information Technology & People 20:352–375
46. Wong KKK-K (2013) 28/05 - Partial Least Squares Structural Equation Modeling (PLS-SEM) Techniques Using SmartPLS. Marketing Bulletin 24
47. Zaleppa P, Dudley A (2020) Ethical, legal and security implications of digital legacies on social media. Lecture Notes in Computer Science 12194. doi: 10.1007/978-3-030-49570-1_29

Part III

Author Index

Author Index