ADAPTIVE GRU-BASED NEURAL NETWORK FRAMEWORK FOR MITIGATING PHISHING ATTACKS IN DIGITAL MICROFINANCE INSTITUTIONS

MATHENGE RICHARD

A THESIS SUBMITTED IN PARTIAL FULFILLMENT FOR AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN INFORMATION TECHNOLOGY IN THE SCHOOL OF PURE AND APPLIED SCIENCES, KIRINYAGA UNIVERSITY.

SEPTEMBER, 2026

Declaration

This thesis is my original work and has not been presented for a degree or any other award in any other University.

Mathenge Nderitu Richard

PA/303/S/21245/23

Signature …  … **Date**: …4-09-2026……

We confirm that the work reported in this thesis was carried out by the candidate under our supervision.

Signature ……  …… **Date**: …07-09-2026…….....

Dr. Catherine Mukunga, PhD

Lecturer, School of Pure and Applied Sciences, Kirinyaga University

Signature …  **Date**: …07-09-2026……....

Dr. Ephantus Mwangi, PhD

Lecturer, School of Pure and Applied Sciences, Kirinyaga University

©

MATHENGE RICHARD

All rights reserved. No part of this thesis may be reproduced, stored in any retrieval system or transmitted in any form or by any means, electronically, mechanically, by photocopying or otherwise, without prior written permission of the author or Kirinyaga University on that behalf.

Dedication

I dedicate this thesis to my family, whose love, unwavering support, encouragement, and sacrifices made this academic journey possible.

Acknowledgement

First and foremost, I give all glory and thanks to God Almighty for His grace, strength, wisdom, and guidance throughout this doctoral journey. His sustaining presence enabled me to persevere through the challenges and successfully complete this work.

The successful completion of this doctoral journey is a testament to the unwavering support, guidance, and encouragement I have received from numerous individuals and institutions. It is with profound gratitude that I acknowledge their contributions.

I extend my deepest appreciation to my supervisors, Dr. Catherine Mukunga and Dr. Ephantus Mwangi. Their expert guidance, insightful critiques, and steadfast encouragement have been invaluable throughout this research process. Their mentorship has not only shaped this thesis but has also profoundly enriched my development as a scholar.

I wish to offer a special and heartfelt acknowledgment to my brother, Dr. Mwehe Mathenge. He has been my mentor, providing unwavering guidance, scholarly advice, and constant encouragement from the very inception of this idea to its conclusion. His wisdom and belief in my abilities were a cornerstone of this achievement.

My most profound gratitude is reserved for my family. To my parents, for their lifelong sacrifices and for instilling in me the value of education. Finally, and most importantly, to my wife and children, whose boundless love, patience, and understanding were my anchor throughout this demanding process. Your sacrifices have been the greatest of all, and this achievement is as much yours as it is mine.

To all who contributed in ways big and small, please accept my sincerest thanks.

Abstract

Microfinance Institutions (MFIs) play a critical role in promoting financial inclusion by providing essential financial services to underserved communities. Nevertheless, growing reliance on digital platforms has made them increasingly vulnerable to sophisticated phishing attacks that threaten operational integrity, compromise sensitive customer information, and erode client trust. Accordingly, this study developed and validated an adaptive Artificial Intelligence (AI)-driven neural network framework to enhance phishing attack detection and mitigation in Nairobi digital microfinance institutions using Gated Recurrent Units (GRUs). The study adopted a mixed-methods research design, integrating a systematic literature review, questionnaires, semi-structured interviews, and experimental deep learning model development, testing, and validation. Primary data were collected from selected Nairobi MFIs, while publicly available phishing datasets comprising URL, email, and network traffic data were used to train and evaluate the proposed GRU model. The framework was implemented using TensorFlow and Keras and validated through controlled experimental evaluation and a three-week pilot deployment. The findings demonstrated that the proposed GRU model achieved a detection accuracy of 94.2%, outperforming comparable Long Short-Term Memory (LSTM) and Convolutional Neural Network (CNN) models while maintaining lower computational overhead suitable for resource-constrained environments. A pilot implementation in a Kenyan MFI further demonstrated the framework's practical feasibility by accurately detecting phishing attempts with a low false-positive rate and supporting near real-time deployment. The study also proposed policy recommendations to strengthen Kenya's cybersecurity legal framework, particularly in AI-driven threat detection, governance, and phishing risk management. The study contributed to knowledge by extending Socio-Technical Systems (STS) Theory through a governance subsystem comprising AI governance, regulatory frameworks, compliance, and institutional oversight, while developing a Phishing Vulnerability–Solution Matrix and an adaptive AI-driven cybersecurity framework tailored to digital MFIs. Overall, the proposed framework provides a scalable, efficient, and context-sensitive approach to strengthening phishing resilience and cybersecurity within digital microfinance institutions.

Abbreviations and Acronyms

AI	Artificial Intelligence
ANOVA	Analysis of Variance
APT	Advanced Persistent Threats
ATO	Account Take Over
BEC	Business Email Compromise
Bi-LSTM	Bidirectional Long Short-Term Memory
CA	Communications Authority
CAAS	Cybersecurity-as-a-service
CBK	Central Bank of Kenya
CEO	Chief Executive Officer
CICIDS	Canadian Institute for Cybersecurity Intrusion Detection System
CMA	Capital Markets Authority
CMCA	Computer Misuse and Cybercrimes Act
CMU	Carnegie Mellon University
CNN	Convolutional Neural Network
COBIT	Control Objectives for Information and Related Technologies
CSF	Critical Success Factor

DMARC	Domain-based Message Authentication, Reporting and Conformance
DPA	Data Protection Act
EAC	East African Community
EU	European Union
FEAT	Fairness, Ethics, Accountability, and Transparency
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GRU	Gated Recurrent Units
HTTPS	HyperText Transfer Protocol Secure
ICT	Information and Communication Technology
IDS	Intrusion Detection Systems
IPS	Intrusion Prevention Systems
ISO/IEC	International Organization for Standardization / International Electrotechnical Commission
KEBS	Kenya Bureau of Standards
KE-CIRT/CC	National Kenya Computer Incident Response Team – Coordination Centre
KPI	Key Performance Indicators
LIME	Local Interpretable Model-agnostic Explanations

LSTM	Long Short-Term Memory
MAS	Monetary Authority of Singapore
MFA	Multi-Factor Authentication
MFI	Microfinance Institutions
ML	Machine Learning
MLP	Multilayer Perceptron
NACOSTI	National Commission for Science, Technology and Innovation
NGO	Non-Governmental Organization
NIST	National Institute of Standards and Technology
CSF	Cybersecurity Framework
NITI	National Institution for Transforming India
RNN	Recurrent Neural Network
ROC-AUC	Receiver Operating Characteristic – Area Under Curve
SACCO	Savings and Credit Cooperative Society
SHAP	SHapley Additive exPlanations
SMS	Short Message Service
TCP	Transmission Control Protocol
URL	Uniform Resource Locator
UTAUT	Unified Theory of Acceptance and Use of Technology

XAI	Explainable Artificial Intelligence

Table of Contents

Declaration…….. ii
Dedication…….. iv
Acknowledgement v
Abstract……. vi
Abbreviations and Acronyms vii
Table of Contents xi
Operational Definition of Key Terms xvii
List of Tables……..... xix
List of Figures…..... xx
CHAPTER ONE INTRODUCTION 1
1.1 Background 1
1.1.1 Cybersecurity Threats in Microfinance Institutions 2
1.1.2 Leveraging Neural Networks in Phishing Detection 3
1.1.3 Towards an Integrated Framework for Countering Cyber Threats 4
1.2 Problem Statement 5
1.3 Objectives 7
1.3.1 General objective 7
1.3.2 Specific Objectives 7
1.4 Research Questions 7
1.5 Justification of the study 8
1.6 Limitations 10
1.7 Delimitations 11
1.8 Organization of the Study 11
1.9 Summary 12
CHAPTER TWO LITERATURE REVIEW 13
2.1 Introduction 13
2.2 Identification of key cybersecurity vulnerabilities in digital MFIs 13
2.2.1 Historical trends of phishing attacks 13
2.2.2 Role of social engineering in phishing threats 15
2.2.3 Emerging AI-driven phishing threats 18
2.2.4 Limitations of Traditional Cybersecurity Solutions in Digital MFIs 19

2.3 Addressing gaps in existing cybersecurity frameworks 21
2.3.1 Collaborative Cybersecurity Models 22
2.3.2 Regulatory Compliance Challenges for MFIs 23
2.3.3 Measuring the Success of Cybersecurity Strategies 24
2.3.4 Global Cybersecurity and AI Governance Frameworks 24
2.3.5 Kenya's Cybersecurity Legal Framework and AI Governance Gaps 27
2.4 Contemporary AI-driven models for detecting phishing attacks 30
2.4.1 Deep Neural Network Models in Cybersecurity 31
2.4.2 Evaluation of Deep Neural Network Models 37
2.4.3 Use of GRUs in Phishing Detection 38
2.4.4 Challenges and Limitations of GRUs in Cybersecurity 38
2.4.5 Explainable Artificial Intelligence (XAI) in AI-driven Phishing Detection 39
2.4.6 Integration of GRUs into MFI Cybersecurity Frameworks 42
2.4.7 Advancing GRUs Integration Through Hybrid and Collaborative Models 43
2.5 Case Studies of Phishing Attacks in Microfinance 46
2.5.1 Analysis of High-Profile Phishing Cases in Microfinance Institutions 46
2.5.2 Analysis of the phishing attack Techniques in the three cases 48
2.5.3 Impact and Defense Mechanisms in all three MFIs 50
2.5.4 Lessons Learned from the Case Studies 51
2.6 Theoretical framework 53
2.6.1 Socio-Technical Systems Theory 53
2.6.2 Institutional Theory 54
2.6.3 Unified Theory of Acceptance and Use of Technology (UTAUT) 56
2.6.4 Justification of the Selected Theories 57
2.6.5 Integration and Application to Research Design 59
2.7 Gaps Identified in the Literature 63
2.8 Conceptual Framework 69
2.9 Summary 72
CHAPTER THREE RESEARCH METHODOLOGY 73
3.1 Introduction 73
3.2 Research philosophy 73
3.3 Research Design 75

3.3.1 Research Strategy 80
3.3.2 Justification for Chosen Design 81
3.4 Research Process 81
3.4.1 Step 1 - Problem formulation and Research objectives 82
3.4.2 Step 2 - Data acquisition 82
3.4.3 Step 3 - Data Processing 83
3.4.4 Step 4 - Model Architecture Design and Training 83
3.4.5 Step 5 - Model evaluation and validation 83
3.5 Population of the Study 84
3.5.1 Sampling Technique 87
3.5.2 Sample Size 88
3.5.3 Inclusion and Exclusion Criteria for Study Participants 90
3.5.4 Justification of Study Area 91
3.6 Data Collection Methods 93
3.6.1 Primary Data Collection Methods 93
3.6.2 Secondary Data Collection Methods 93
3.6.3 Development of Data Collection Tools 94
3.6.4 Questionnaire and Interview Guide Instrument Validity 95
3.7 Data Analysis 97
3.7.1 Qualitative Analysis 97
3.7.2 Quantitative Analysis 97
3.7.3 Data Analysis Software 99
3.7.4 Data Integration 99
3.8 Dataset Strategy 99
3.8.1 Dataset selection 99
3.8.2 Dataset Justification and Relevance 101
3.9 Ethical Considerations 102
3.9.1 Informed Consent 102
3.9.2 Confidentiality and Data Privacy 102
3.9.3 Ethical Approval 103
3.9.4 Addressing Ethical Challenges in Real-Time Model Testing 103
3.10 Summary 103

CHAPTER FOUR RESULTS AND DISCUSSION 105
4.1 Introduction 105
4.2 Cybersecurity Vulnerabilities and Phishing Threats in Digital Microfinance Institutions ..105
4.2.1 Data Sources and Analytical Approach 106
4.2.2 Rationale for Focusing on Phishing in MFIs 107
4.3 Objective 1 – mapping cybersecurity vulnerabilities facing digital MFIs that can be addressed through AI-driven solutions 108
4.3.1 MFI Cyber Threat Landscape 108
4.3.2 AI-Driven Mitigation of Phishing and Related Threats 110
4.4 Objective 2 – Cybersecurity and AI Policy Frameworks 112
4.4.1 Global Cybersecurity Frameworks Overview 113
4.4.2 Legal and Policy Gaps in Kenya's Cybersecurity Framework 115
4.4.3 Proposed Amendments to the CMCA (2018) 119
4.5 Objective 3 – GRU-Based Phishing Detection Model 125
4.5.1 GRU Justification 125
4.5.2 Dataset Preparation 126
4.5.3 Model Architecture 127
4.5.4 Model Performance and Comparison 130
4.5.5 Challenges in Developing and Testing the GRU Model 133
4.5.6 Model Training and Evaluation 134
4.6 Objective 4 – Model Validation with MFI Data 139
4.6.1 Pilot Setup and Validation Strategy 139
4.6.2 Validation Results 140
4.6.3 Ethical Implications of the Findings 142
4.7 Research Contributions Per Objective 143
4.7.1 Findings Emerging from the Data 147
4.8 Critical Synthesis and Strategic Interventions 149
4.8.1 Study Synthesis 149
4.8.2 Strategic Interventions 150
4.9 Summary 152
4.10 DISCUSSION 153
4.11 Analysis by Research Objectives 155

4.11.1 Objective 1 – Cybersecurity Vulnerabilities and Phishing in MFIs 155
4.11.2 Objective 2 – Governance and Policy Frameworks 157
4.11.3 Objective 3 – GRU-Based Phishing Detection Model Development 159
4.11.4 Objective 4 - Model Validation 161
4.12 Cross-Objective Insights and Integration 163
4.12.1 Linking Vulnerabilities, Governance, and Validation 164
4.12.2 Emergent Thematic Insights 164
4.12.3 Contribution to Knowledge and Practice 165
4.13 Theoretical and Practical Implications 166
4.13.1 Theoretical Implications 166
4.13.2 Practical Implications 168
4.13.3 Bridging Theory and Practice 168
4.14 Policy and Regulatory Implications 169
4.14.1 Global and Regional Perspectives 169
4.14.2 National-Level Gaps and Opportunities 171
4.14.3 Risk-Based and Adaptive Governance Lens 172
4.14.4 Feasibility and Institutional Capacity 172
4.15 Practical Interventions for MFIs 173
4.16 Summary 175
CHAPTER FIVE SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS 177
5.1 Introduction 177
5.2 Integrated Conclusions 177
5.3 Contributions of the Study 179
5.3.1 Contribution to Knowledge 180
5.4 Practical and Policy Recommendations 182
5.5 Recommended Amendments to the CMCA (2018) 183
5.6 Future Research Directions 185
REFERENCES 187
APPENDICES 199
Appendix 1: Questionnaire 199
Appendix 2: Interview Schedule 209
Appendix 3: Public Datasets for GRU-Based Phishing Detection 212

Appendix 4: List of Licensed and authorized Saccos Societies in Nairobi, Kenya 213
Appendix 5: University Introduction Letter 255
Appendix 6: NACOSTI Research Permit 256
Appendix 7: Code Repository 257
Appendix 8: Certificate of Acceptance (JCS) 261
Appendix 9: Certificate of Acceptance (IJNSA) 262

Operational Definition of Key Terms

Adaptive Cybersecurity: This refers to the capability of a cybersecurity system to continuously learn from emerging threats and dynamically adjust its detection and response mechanisms. In this study, it describes the proposed AI-driven framework that continually improves phishing detection using new threat data.

Artificial Intelligence (AI): This refers to computational techniques that enable computer systems to perform tasks requiring human-like intelligence, including pattern recognition, learning, and decision-making. In this study, AI specifically refers to deep learning techniques employed for phishing detection and mitigation in digital microfinance institutions.

Cybersecurity Governance: This refers to the policies, regulatory frameworks, institutional structures, and oversight mechanisms that guide the management of cybersecurity risks and ensure compliance with legal and organizational requirements. In this study, it encompasses governance practices supporting the adoption of AI-driven phishing detection within digital MFIs.

Deep Learning: Deep learning refers to a subset of machine learning that utilizes multi-layered artificial neural networks to automatically learn complex representations from data. In this study, deep learning forms the basis for developing and evaluating the GRU-based phishing detection model.

Digital Microfinance Institution (MFI): This refers to a microfinance institution that delivers financial services through digital platforms, including mobile applications, internet-based services, and electronic payment systems. This study focuses on Nairobi's digital MFIs that provide technology-enabled financial services.

Gated Recurrent Unit (GRU): A Gated Recurrent Unit (GRU) is a deep learning model designed to learn patterns from sequential data. In this study, the GRU serves as the core deep learning model for detecting phishing attacks in emails, URLs, and related digital communications.

Phishing: This is a form of cyberattack in which attackers impersonate legitimate entities through fraudulent emails, websites, SMS messages, or other digital communication channels to deceive users into disclosing sensitive information or performing unauthorized actions.

Phishing Detection: Refers to the process of automatically identifying and classifying digital communications as phishing or legitimate using artificial intelligence techniques. In this study, phishing detection is achieved through the proposed GRU-based framework.

Phishing Vulnerability–Solution Matrix: It refers to the framework developed in this study that systematically maps identified phishing vulnerabilities in digital microfinance institutions to corresponding AI-driven mitigation strategies. The matrix provides a structured approach for prioritizing phishing risk management and strengthening institutional cybersecurity resilience.

Socio-Technical Systems (STS) Theory: This theory explains organizational performance through the interaction of social and technical subsystems. In this study, the theory is extended by incorporating a governance subsystem comprising AI governance, regulatory frameworks, compliance, and institutional oversight.

List of Tables

Table 2.1: Cooperative Overview of Global Cybersecurity and AI Governance Frameworks..... 26

Table 2.2 Kenya's Cybersecurity Legal Frameworks and Gaps for AI Governance 28

Table 2.3 Comparative Characteristics of RNN, LSTM, GRU, and Transformer Architectures in Phishing Detection Applications..... 34

Table 2.4 Summary of Case Studies 49

Table 2.5 Operationalization of the Integrated Theoretical Framework 61

Table 2.6 Summary of gaps in the literature on phishing attacks on digital MFIs 66

Table 3.1: Research designs and research questions addressed in the different chapters of the thesis 77

Table 3.2 Distribution of Retrieved Studies by Source..... 85

Table 4.1 Global Cybersecurity and AI Governance Frameworks 114

Table 4.2 Kenyan Cybersecurity Legal Frameworks and AI Gaps.....116

Table 4.3 Policy Misalignments in MFI Cybersecurity Governance.....118

Table 4.4 Proposed Amendments to CMCA (2018) for AI Readiness 121

Table 4.5 GRU Model Architecture 128

Table 4.6 Model Performance Comparison 131

Table 4.7 Performance Metrics from Pilot Validation 141

Table 4.8 Summary of Contributions per Objective 146

Table 4.9 Proposed Amendments with Feasibility Analysis..... 147

Table 4.10 Global Cybersecurity and AI Governance Frameworks 170

Table 5.1 Proposed Amendments to CMCA (2018) 183

List of Figures

Figure 2.1Integrated Theoretical Framework for the Study 60

Figure 2.2 Conceptual Framework for AI-Driven Phishing Detection..... 72

Figure 3.1: Step-by-step Research Process..... 82

Figure 4.1 GRU phishing detection model pipeline 127

Figure 4.2 Illustrative code snippet..... 130

Figure 4.3 Training/validation learning curves. 137

CHAPTER 1

INTRODUCTION

1.1 Background

Microfinance institutions (MFIs) play a vital role as a financial lifeline to the underprivileged populations, serving the basic needs of small business owners, low-income individuals, and rural communities. The institutions are instrumental in promoting financial inclusion and economic growth in sectors that are often overlooked by traditional financial services. However, the digital revolution that has widened the scope of MFIs also subjects them to a broader range of cybersecurity threats. Among them, phishing attacks pose a particularly pernicious threat, with the potential to undermine the trust and security that MFIs are trying to build (Ali et al., 2024).

Phishing attacks, in which cybercriminals trick people into providing sensitive information, have become increasingly sophisticated. The consequences of such attacks on MFIs can be devastating, leading to data breaches, loss of funds, and extensive loss of client trust (Ali et al., 2024; Javaheri et al., 2023). However, most MFIs lack the resources and expertise to implement advanced cybersecurity measures, which makes them appealing targets for such activities (Non, 2023). Traditional cybersecurity methods, despite their fundamental nature, cannot keep pace with the rapidly evolving tactics used by cybercriminals. The use of firewalls, intrusion detection systems, and antivirus software, although vital, is often insufficient to counter the complex and adaptive nature of modern phishing attacks (Kiganda, 2022).

The rise of Artificial Intelligence (AI) and Machine Learning (ML) offers a ray of hope in this challenging environment. By leveraging these technologies, it is possible to create intelligent and adaptable systems that can identify and mitigate phishing attacks in real time (Shah, 2021). This study aims to explore how advanced neural network architectures, namely the Gated Recurrent

Units (GRUs), can be harnessed to enhance phishing detection within digital microfinance institutions. The selection of GRUs was informed by previous studies demonstrating their effectiveness in sequential text classification and phishing detection (Roy et al., 2022).

1.1.1 Cybersecurity Threats in Microfinance Institutions

Microfinance operations are digital in nature; therefore, MFIs will always be susceptible to various cyber threats. These institutions handle highly sensitive financial information, including transaction records, clients' personal information, and financial histories. The integrity and reputation of MFIs are not the only reasons why such data security is paramount; it is also for the safety of individual clients. An attack can lead to major financial losses and loss of trust, a factor that can be particularly harmful in societies where MFIs are a primary means of accessing financial services (Javaheri, 2023).

The cybersecurity landscape has become more complex over the last few years, and its threats continue to advance in both sophistication and scale. Phishing attacks have become more targeted and cunning, often using social engineering mechanisms, which manipulate human psychology to bypass traditional security measures (Camacho, 2024). These are not isolated cases; they represent a broader trend in which cybercriminals continually innovate, using new tools and techniques to compromise systems. Therefore, given the dynamic nature of the threat environment, MFIs will require cybersecurity strategies that are not only reactive but also predictive and adaptive.

Conventional cybersecurity models, though fundamental, tend to rely on static defenses such as firewalls and antivirus software, which are primarily designed to block known threats. Such systems, however, are often outpaced by the dynamism of modern cyber threats, which can rapidly evolve to exploit new vulnerabilities (Nicholls, 2021). This weakness is especially acute in the

MFI scenario, where resource scarcity hampers the continuous updating and management of advanced cybersecurity infrastructure.

1.1.2 Leveraging Neural Networks in Phishing Detection

The shortcomings of traditional cybersecurity methods underscore the need for advanced approaches that can address the unique threats faced by Microfinance Institutions (MFIs). Neural networks are a class of machine learning models designed to learn complex patterns from data through interconnected processing units (neurons). In this respect, neural networks, especially Gated Recurrent Units (GRUs), come into play. GRUs are a type of recurrent neural network (RNN) specifically designed to process sequential data by retaining relevant information over time through gating mechanisms (Salem, 2021). Their ability to process sequential data makes them particularly well-suited to detecting phishing attacks, which are often characterized by patterns in emails, uniform resource locators (URLs), or user behavior over time (Roy et al., 2022).

GRUs perform simplified sequence processing that uses fewer parameters while retaining the ability to capture crucial patterns. They excel in areas where computational facilities are limited, thus a perfect option for MFIs that may not have access to elaborate IT infrastructure. Besides, although their architecture is relatively simple, GRUs are effective at capturing long-term dependencies, which helps them discover the subtle, evolving patterns characteristic of phishing activities (Saha et al., 2020).

This research aims to examine the application of an AI-driven adaptive framework leveraging a GRU-based model to develop an efficient and effective solution for phishing detection in MFIs. The goal is to develop a structure that not only improves detection accuracy but also adapts to new threats, providing a strong defense against an ever-changing cyber threat landscape.

1.1.3 Towards an Integrated Framework for Countering Cyber Threats

Given the diverse nature of cyber threats and the sensitivity of the financial data that microfinance institutions (MFIs) handle, there is a pressing need to develop a comprehensive cybersecurity framework to effectively address these challenges. A built-in framework would track network activity, detect anomalies, and automatically initiate countermeasures to reduce risk before it causes significant damage. This methodology helps to preempt threat hunting and continuous adaptation through feedback loops and event data analytics (Mwavali, 2024).

Such a framework should incorporate multiple layers of defense mechanisms and technologies to create a strong cybersecurity posture. Rather than relying on isolated security controls, the proposed integrated framework combines threat intelligence, AI-driven analytics, continuous monitoring, and automated response mechanisms into a coordinated cybersecurity framework that supports the prevention, detection, and mitigation of cyber threats. By integrating these components, the framework enables the continuous exchange of threat information, thereby improving situational awareness and facilitating timely responses to evolving phishing attacks and other cybersecurity threats affecting digital MFIs.

Within this integrated framework, artificial intelligence serves as the analytical component that strengthens threat detection by analyzing large volumes of security events, user behavior, and network activities to identify suspicious patterns associated with phishing attacks. The AI capability, implemented using a Gated Recurrent Unit (GRU) neural network model, complements existing cybersecurity controls by providing adaptive threat detection while supporting automated decision-making and continuous learning from emerging attack patterns (Mughaid et al., 2022).

The integration of machine learning into existing threat detection systems enhances their effectiveness, providing faster, more accurate detection with a low false-positive rate. This is of great importance to MFIs, since reducing false positives allows security personnel to focus on genuine threats, thereby maximizing overall security performance. Moreover, machine learning models can continuously learn and adapt to new threats, making detection systems more effective in a constantly changing threat landscape (Gautam, 2021). Consequently, the integrated framework enhances the cybersecurity resilience of digital MFIs by combining AI-driven threat detection with continuous monitoring, coordinated incident response, and adaptive security mechanisms, enabling institutions to respond proactively to evolving cyber threats rather than relying solely on reactive security measures.

1.2 Problem Statement

Microfinance Institutions (MFIs) play a significant role in providing financial services to underprivileged communities, contributing to economic development and financial inclusion (Adnan et al., 2021). However, the rapid digitalization of financial services has increased their exposure to sophisticated phishing attacks. Phishing remains one of the most persistent cyber threats because it exploits technological vulnerabilities and human behavior, resulting in financial losses and erosion of customer trust (Joseph, 2023; Webb & Sallos, 2024). The Anti-Phishing Working Group (APWG, 2024) reported millions of phishing campaigns annually, demonstrating the persistence of the threat. In Kenya, digital MFIs face increasing exposure to phishing attacks due to constrained financial resources, limited technical expertise, and inadequate cybersecurity infrastructure (Webb & Sallos, 2024). The Communications Authority of Kenya also reported

continued growth in cyber threat events targeting critical sectors, including financial services (Communications Authority of Kenya [CA], 2024).

Although existing cybersecurity measures and frameworks provide structured approaches to risk management and threat detection, their reliance on predefined controls and periodic assessments can limit their effectiveness against rapidly evolving phishing attacks (Taofeek, 2024; Khaleefah et al., 2024). Consequently, digital MFIs continue to face phishing-related incidents that compromise operations and client confidence (Ali et al., 2024). GRUs have demonstrated applicability in sequential data analysis and have also been applied to phishing detection, including phishing URL classification, where their ability to capture sequential patterns and operate with comparatively lower computational requirements presents potential advantages for resource-constrained environments (Roy et al., 2022; Kar et al., 2024). However, the existing application of GRUs has not adequately addressed the integration of adaptive phishing detection with vulnerability assessment, cybersecurity governance, and the operational constraints of resource-constrained digital MFIs. Previous studies have examined phishing detection and cybersecurity in financial institutions, but limited research has focused on integrating vulnerability assessment, governance, and adaptive detection within resource-constrained digital MFI environments (Almujahid et al., 2024; Ali et al., 2024). In particular, limited evidence exists on an integrated adaptive GRU-based framework for phishing detection within the digital MFI context in Kenya. This research gap established the need for the current study.

1.3 Objectives

1.3.1 General objective

To design an AI-driven adaptive cybersecurity framework for mitigating phishing attacks in digital microfinance institutions.

1.3.2 Specific Objectives

a) To map cybersecurity vulnerabilities facing digital microfinance Institutions that can be addressed through AI-driven solutions.

b) To assess the effectiveness of existing cybersecurity frameworks in mitigating phishing threats within digital microfinance institutions.

c) To develop a neural network AI-driven adaptive framework, using GRU neural network models, for improving phishing threat detection in digital microfinance Institutions.

d) To validate the effectiveness of the developed AI-driven adaptive framework in detecting and mitigating phishing threats within digital microfinance Institutions.

1.4 Research Questions

a) How can the key cybersecurity vulnerabilities in digital microfinance Institutions be identified and mitigated using AI-driven solutions?

b) How effective are existing cybersecurity frameworks and policies in mitigating evolving cybersecurity threats facing microfinance Institutions?

c) How can GRU neural network models be optimized to accurately detect, prevent, and predict phishing threats within digital microfinance Institutions?

d) How effective is the newly developed AI-driven framework in improving the detection and mitigation of phishing threats compared to existing methods within digital microfinance Institutions?

1.5 Justification of the study

This research was crucial as it addressed the evolving cybersecurity challenges facing Microfinance Institutions (MFIs), which are essential to delivering financial services to underserved populations. The growing sophistication of phishing attacks, together with the increasing adoption of digital financial services, exposed resource-constrained MFIs to heightened cybersecurity risks, leading to financial losses and reduced client trust (Joseph, 2023; Webb & Sallos, 2024). The scale of the phishing problem was evident at the global level, with the Anti-Phishing Working Group (APWG) reporting 963,994 phishing attacks in the first quarter of 2024 and a further 877,536 attacks in the second quarter of the same year (APWG, 2024). These figures indicated that phishing remained a persistent and substantial cybersecurity threat, with reported attacks remaining at a high level across successive quarters. In Kenya, the Communications Authority of Kenya (CA) reported 971,440,345 cyber threat events between January and March 2024 and 1,131,696,878 events between April and June 2024, representing a 16.5% increase between the two quarters (CA, 2024). The reported cyber threat activity included system attacks, malware attacks, brute-force attacks, distributed denial-of-service attacks, web application attacks, and mobile application attacks, demonstrating the breadth of the cybersecurity threat environment in Kenya. These figures demonstrated that the cybersecurity problem extended beyond isolated incidents and occurred at a scale capable of affecting the continuity, security, and trust associated with digital financial services. The increasing complexity of phishing, as well as the limited

cybersecurity resources of MFIs, underscored the need for more adaptive cybersecurity approaches capable of responding to rapidly evolving cyber threats.

Although existing cybersecurity measures, including the National Institute of Standards and Technology (NIST) Cybersecurity Framework and ISO/IEC 27001, provided structured guidance for cybersecurity governance and risk management, they had not fully addressed the unique operational and resource constraints of digital MFIs in responding to evolving phishing attacks (Taofeek, 2024; Khaleefah et al., 2024). This policy and practice gap provided a strong justification for undertaking this study.

This study utilized artificial intelligence, specifically the Gated Recurrent Unit (GRU) neural network, to improve the detection and mitigation of phishing attacks. An AI-based adaptive framework tailored to MFIs was developed to strengthen their cybersecurity resilience and protect their operations and clients' confidential financial data.

Furthermore, previous studies reported limited research on integrated AI-driven cybersecurity frameworks designed specifically for digital MFIs (Almujahid et al., 2024). The combination of the documented global scale of phishing, the increasing volume of cyber threats reported in Kenya, and the limited availability of MFI-specific AI-driven cybersecurity frameworks demonstrated both the magnitude of the problem and the gap requiring further investigation. If this gap had remained unaddressed, digital MFIs would have continued to experience phishing-related financial losses, operational disruptions, regulatory compliance challenges, and declining customer confidence, ultimately affecting the sustainability of digital financial services. The findings of this study provided practical evidence to support digital MFIs, policymakers, cybersecurity practitioners, and researchers in strengthening cybersecurity resilience against evolving phishing attacks.

1.6 Limitations

The relevance of this study to microfinance institutions (MFIs) with more developed digital platforms may be limited by purposive sampling. Nairobi County was selected because it hosts a high concentration of digitally enabled MFIs, making it an appropriate setting for evaluating the proposed AI-driven cybersecurity framework. Results were contextualized within Nairobi to alleviate this constraint, with the understanding that they might not be comprehensive for MFIs with varying resources or less reliance on digital technology. Accordingly, the findings should be interpreted with caution when generalizing to MFIs operating in other counties with different levels of digitalization and cybersecurity maturity.

Despite the impressive performance of the GRU model, it may still have trouble with identifying complex and advanced phishing attacks. Restricted access to large phishing datasets during training might affect the model's accuracy. In addition, the pilot validation was conducted using data from a single operational MFI because access to cybersecurity data from multiple institutions was constrained by confidentiality and organizational data-sharing restrictions. These limitations are discussed in this study to provide a comprehensive assessment of the model's competence.

Since cybersecurity data is sensitive and its confidentiality needs are high, the accessibility and disclosure of some data may be restricted in some organizations. Sensitive results were anonymized to mitigate this risk and reduce the likelihood of exposing participants. Appropriate ethical procedures and confidentiality measures were also implemented throughout data collection, storage, and analysis to protect both participant and institutional information.

1.7 Delimitations

This study focused solely on phishing attacks targeting MFIs in Nairobi, excluding malware and other cybersecurity threats. This delimitation reduces the scope of the investigation, enabling scrutiny of phishing vulnerabilities in a specific urban digital environment.

1.8 Organization of the Study

This thesis is organized into five chapters.

Chapter One introduces the study by presenting the background, problem statement, research objectives, research questions, significance of the study, scope, limitations, definition of key terms, and the theoretical and conceptual foundations underpinning the research.

Chapter Two presents the literature review, examining existing studies on phishing attacks, artificial intelligence, cybersecurity in digital microfinance institutions, governance and policy frameworks, and neural network-based phishing detection. The chapter also identifies the research gaps that informed the development of the proposed framework.

Chapter Three describes the research methodology adopted in the study. It outlines the research paradigm, research design, target population, sampling procedures, data collection methods, experimental design, data analysis techniques, ethical considerations, and the procedures followed in developing and validating the proposed GRU-based framework.

Chapter Four presents the findings and discussion. It reports the results for each research objective, including the identified cybersecurity vulnerabilities, governance and policy analysis,

development and evaluation of the GRU-based phishing detection framework, pilot validation, and the study's contributions to knowledge.

Chapter Five presents the summary of the findings, conclusions, contributions to knowledge, study limitations, recommendations for policy and practice, and suggestions for future research. The chapter concludes by highlighting the implications of the proposed AI-driven adaptive cybersecurity framework for enhancing phishing resilience in digital microfinance institutions.

1.9 Summary

The research addressed one of the critical issues faced by microfinance institutions, i.e., the need for advanced, flexible cybersecurity frameworks capable of protecting against increasingly sophisticated phishing attacks. The study aimed to leverage gated recurrent unit (GRU) neural networks to develop a robust, scalable framework that significantly improved the security posture of MFIs. The study provided the roadmap for improving the cybersecurity resiliency of these digital microfinance organizations through comprehensive analysis, implementation, and validation.

CHAPTER 2

CHAPTER 3

LITERATURE REVIEW

3.1 Introduction

The chapter provides an in-depth review of the literature on cybersecurity concerns and developments in microfinance institutions (MFIs), with particular attention to the growing threat of phishing. The development of phishing attacks, including AI-based ones, is discussed, along with their effects on the operational, financial, and reputational stability of MFIs. The chapter explores the potential of AI-based solutions, such as Gated Recurrent Units (GRUs), to address these problems, questions the weaknesses of traditional cybersecurity practices, and identifies gaps in current frameworks. It has synthesized previous studies and case studies to emphasize practical findings that improve MFI detection and phishing resistance, given resource constraints.

3.2 Identification of key cybersecurity vulnerabilities in digital MFIs

3.2.1 Historical trends of phishing attacks

In Over the past decade, phishing attacks have evolved significantly in both their sophistication and their choice of targets. Early phishing campaigns primarily targeted large commercial banks through mass email attacks designed to obtain customers' login credentials and financial information. As digital financial services became more widely adopted, cybercriminals shifted their attention to microfinance institutions (MFIs), recognizing that many of these institutions operated with limited cybersecurity resources, inadequate technical capacity, and less mature security governance structures ((International Criminal Police Organization [INTERPOL], 2024); Alkhalil et al., 2021). This transition reflected a broader change in the cyber threat landscape,

where attackers increasingly exploited institutional weaknesses rather than relying solely on technical system vulnerabilities.

Research has shown that phishing attacks against financial institutions have become increasingly targeted and personalized. Ahmad et al. (2021) reported that cybercriminals adopted techniques such as spear-phishing, smishing, and whaling to exploit weaknesses in organizational processes and human behavior. Unlike traditional phishing campaigns that targeted large numbers of users indiscriminately, these techniques focused on specific individuals or organizational roles, thereby increasing the likelihood of successful compromise. This evolution demonstrated that phishing had become a strategic attack vector capable of bypassing conventional security controls through manipulation of trust and organizational communication channels (Ahmad et al., 2021; Zhang et al., 2023).

Individual case studies further illustrated this growing trend. Smith and Adewale (2019) reported a spear-phishing attack against an African MFI in 2018 that resulted in the theft of executive credentials valued at more than USD 500,000. Similarly, Nguyen et al. (2020) documented a whaling attack in which the Chief Executive Officer of a microfinance institution in Southeast Asia authorized a fraudulent wire transfer of approximately USD 1.2 million after responding to a carefully crafted phishing campaign. Although these incidents represented individual cases, they reflected patterns that have been consistently reported across the literature, namely that phishing attacks increasingly exploited organizational authority, employee trust, and weaknesses in internal verification processes rather than purely technological vulnerabilities (Alkhalil et al., 2021; Ahmad et al., 2021). These cases should therefore be interpreted as illustrative examples of broader trends rather than evidence of the prevalence of phishing attacks across all MFIs.

The reviewed literature consistently suggested that the vulnerability of MFIs extended beyond limitations in information technology infrastructure. Resource constraints, limited cybersecurity investment, inadequate staff awareness, and evolving digital service delivery models collectively increased the exposure of MFIs to phishing attacks (Nicholls et al., 2021). While existing studies acknowledged these vulnerabilities, most examined them independently, with limited attention given to how technical, organizational, and human factors interacted to influence phishing resilience within digital MFIs. Consequently, the literature provided only a fragmented understanding of the cybersecurity challenges facing MFIs, leaving an important gap regarding the development of integrated and adaptive cybersecurity approaches capable of addressing these interrelated vulnerabilities.

3.2.2 Role of social engineering in phishing threats

Phishing Social engineering has become one of the most effective techniques used in phishing attacks against microfinance institutions (MFIs). Unlike attacks that exploit technical weaknesses in information systems, social engineering relies on manipulating human behavior to gain unauthorized access to sensitive information. Attackers typically impersonate trusted individuals or organizations and use communication channels such as emails, text messages, and telephone calls to deceive employees into disclosing confidential information or performing actions that compromise organizational security (Duarte et al., 2021). The increasing digitization of financial services has expanded these opportunities, allowing cybercriminals to combine technical deception with psychological manipulation in ways that are often difficult to detect using conventional security controls.

As phishing attacks have evolved, cybercriminals have adopted more targeted approaches that exploit different forms of communication and varying levels of organizational authority. Ahmad et al. (2021) identified spear-phishing, smishing, and voice phishing (vishing) as some of the most common techniques used to compromise financial institutions. Vishing refers to phishing attacks conducted through telephone or Voice-over-Internet Protocol (VoIP) calls, where attackers impersonate trusted individuals or organizations to persuade victims to disclose confidential information or perform unauthorized actions. Rather than targeting large groups of users indiscriminately, these attacks were designed to exploit trust relationships within organizations by impersonating supervisors, colleagues, financial partners, or technical support personnel. This shift reflected a significant change in the nature of phishing attacks, where success increasingly depended on influencing human judgement instead of bypassing technical security mechanisms. Consequently, employee awareness and organizational security practices became just as important as technological safeguards in reducing phishing-related risks (Alkhalil et al., 2021).

The growing use of artificial intelligence (AI) has further changed the phishing landscape. AI-enabled tools have made it possible for attackers to generate convincing phishing messages that closely resemble legitimate organizational communication, making fraudulent emails increasingly difficult to distinguish from genuine correspondence. Joseph (2023) reported a phishing incident involving a South American MFI in which AI-generated emails contributed to a significant security breach. Although this represented an individual case, it reflected a broader concern highlighted in recent studies that AI has increased the speed, scale, and sophistication of phishing campaigns, thereby creating new challenges for financial institutions with limited cybersecurity capabilities (Zhang et al., 2023). While these studies have demonstrated how AI has strengthened phishing attacks, comparatively less attention has been given to the development of adaptive AI-

driven defensive mechanisms tailored to resource-constrained financial institutions. These developments suggest that traditional rule-based detection approaches may no longer be sufficient to respond effectively to rapidly evolving phishing techniques.

Evidence from previous studies also demonstrated that social engineering attacks continued to exploit weaknesses in organizational processes. Adebayo (2022) documented a case in which employees of a Nigerian microfinance institution unknowingly reset organizational passwords during a vishing attack, allowing unauthorized access to institutional systems. While this example illustrated the operational impact of social engineering, similar findings across the literature suggested that human error remained one of the most exploited vulnerabilities in phishing attacks targeting financial institutions (Duarte et al., 2021; Ahmad et al., 2021). These observations emphasized that effective phishing prevention required more than technical security controls. Strong verification procedures, continuous staff awareness programs, and a security-conscious organizational culture were equally important in reducing the likelihood of successful attacks.

The impact of social engineering has also been evident within the African financial sector. Mwangi and Otieno (2021) reported that a Kenyan MFI approved fraudulent loans valued at approximately USD 400,000 following a phishing attack. Likewise, Nkosi (2023) observed that a South African microfinance institution experienced a decline in customer loyalty after confidential client information was exposed during a phishing incident. Smith et al. (2023) further reported that a majority of surveyed MFIs affected by phishing attacks experienced operational disruptions, increased compliance obligations, and additional recovery costs. Although these studies reflected different institutional contexts, they consistently demonstrated that the consequences of phishing extended beyond immediate financial losses to include reputational damage, service interruptions,

regulatory exposure, and declining customer confidence. Collectively, these findings demonstrated that phishing has evolved beyond a technical security problem to become an organizational risk affecting operational continuity, regulatory compliance, institutional reputation, and customer confidence. This pattern suggested that phishing had become both a cybersecurity challenge and an organizational resilience issue for digital MFIs.

Despite the growing body of research on phishing and social engineering, important gaps remain. Much of the existing literature has concentrated on describing attack techniques and documenting their immediate effects, while comparatively little attention has been given to how technical controls, employee awareness, organizational governance, and adaptive artificial intelligence can be integrated to strengthen phishing resilience in digital MFIs. Similarly, limited empirical evidence exists on the long-term effectiveness of continuous cybersecurity awareness programs within resource-constrained financial institutions. Furthermore, most existing studies have evaluated phishing detection techniques independently, with limited consideration of how technical detection models can be integrated with governance and organizational security practices in operational MFI environments. These gaps provided the motivation for this study to develop an integrated AI-driven cybersecurity framework that combined phishing vulnerability assessment, governance and policy considerations, and an adaptive GRU-based phishing detection model to enhance the cybersecurity resilience of digital MFIs.

3.2.3 Emerging AI-driven phishing threats

During this modern age of artificial intelligence, new phishing trends such as deep-fake technology and artificial intelligence-driven phishing are new threats to the microfinance institutions (MFIs). In 2022, an Indian microfinance institution fell victim to a deepfake phishing call that resulted in

a fraudulent transfer of $300,000, in which the attacker impersonated the chief executive officer (Patel & Gupta, 2023). Equally, the implementation of artificial intelligence in phishing attacks enables cybercriminals to carry out highly personalized attacks on a larger scale. An African MFI in 2023 reported that phishing attempts using AI-generated emails increased by 40% (Musa & Amina, 2023). Moreover, the use of multiple channels in phishing, such as email, SMS, and social media, hinders MFIs' ability to identify and prevent potential access points. Given the emerging risk of mobile phishing, MFIs should implement AI-powered security interventions to protect against more advanced attacks (Pattnaik et al., 2024). Although scholarship on AI-powered attacks is growing, the current literature ignores the technological constraints that MFIs face when implementing AI-based security measures. Further investigation into how AI-based solutions can be adapted to the specific requirements of such institutions is required to ensure that MFIs with limited technical capabilities can protect themselves against phishing attacks without incurring significant infrastructure costs.

3.2.4 Limitations of Traditional Cybersecurity Solutions in Digital MFIs

The cybersecurity weaknesses that microfinance institutions (MFIs) face stem from resource limitations, outdated technologies, and a lack of expertise in risk management. Despite their strong position to achieve financial inclusion, these institutions are often ill-prepared to cope with modern threats such as AI-driven phishing attacks, advanced social engineering, and ransomware (Alkhalil et al., 2021). Unlike large financial institutions, MFIs lack sufficient capital and technical expertise to maintain a strong cybersecurity defense system and are consequently more vulnerable to human error and targeted attacks (Amy et al., 2021). Employees, who are usually poorly trained in cybersecurity, often become victims of a phishing attack, leading to data breaches, massive

financial losses, and legal regulatory penalties (Ali et al., 2023). Even though classical cybersecurity models can help mitigate threats, they usually fall short of curbing the ever-changing threat landscape that MFIs face in the modern digital environment.

Conventional security solutions, such as firewalls, Intrusion Detection Systems (IDS), Intrusion Prevention Systems (IPS), and antivirus software, remain in use across sectors, including MFIs. However, most of these technologies are designed to identify known threats through signature-based approaches, and thus only malware and phishing attacks with predefined patterns can be detected (Alanezi, M. 2021). All the MFIs are therefore susceptible to zero-day vulnerabilities, advanced persistent threats (APTs), and new social engineering tricks, due to the reactive position of these tools (Mallick et al., 2024). Traditional solutions often do not protect encrypted traffic, and multi-vector attacks are used in combination across various ingress points in parallel, including email, SMS, and social media services (Xu et al., 2020).

Limitations of traditional measures, especially in microfinance institutions (MFIs), are distinct, and they face challenges in keeping their technology up to date due to financial constraints. Most MFIs still use outdated systems, making them vulnerable to advanced phishing attacks that conventional detection systems cannot detect (Tian et al., 2021). These vulnerabilities allow attackers to organize spear-phishing campaigns or create AI-generated phishing emails that appear legitimate and can trick even cautious workers. Furthermore, a shortage of cybersecurity expertise on the internal level makes MFIs unprepared to deal with such threats, thus increasing their chances of unauthorized access, data breaches, and loss of money (Shafiq et al., 2020).

The National Institute of Standards and Technology (NIST) Cybersecurity Framework is a detailed assortment of guidelines to help industries and institutions manage the risks of cybersecurity. The

framework supports the improvement of the capability of an institution to evaluate, identify, and respond to cybersecurity threats. Another relevant model is the ISO/IEC 27001, which is well known due to its information-security management framework that has a systematic method of risk management. These models focus on risk management, continuous monitoring, and incident response as core components of developing cybersecurity resilience (Malatji, 2023). However, MFIs cannot adapt these frameworks due to their complexity and resource requirements (Ali et al., 2023). Whereas larger financial institutions may deploy these frameworks in their entirety, MFIs need a simplified model that prioritizes core security measures to avoid exceeding their limited budgets (Ibrahimnur, 2023).

To make these frameworks more relevant to MFIs, one should place greater emphasis on the following key areas: encryption, authentication protocols, and employee training. In a phased implementation approach, MFIs can start with basic cybersecurity measures, such as multi-factor authentication (MFA) and regular system updates, and then advance to more sophisticated methods, such as behavioral analytics and continuous threat monitoring (Xu et al., 2020). For example, the Zero Trust Architecture, which emphasizes checking every stage of digital interactions, can be simplified for MFIs to focus on key aspects such as access control and user identity verification (Gupta et al., 2021). This risk-based model allows MFIs to strengthen their cybersecurity while staying within financial constraints.

3.3 Addressing gaps in existing cybersecurity frameworks

One of the significant shortcomings of modern cybersecurity frameworks is their failure to effectively counter AI-driven threats and constantly evolving social engineering tactics. Attackers use AI to craft highly realistic phishing emails and conduct behavioral analysis of their potential

victims, further diminishing the effectiveness of traditional defense mechanisms (Tian et al., 2021). For example, AI-powered phishing attacks create custom phishing messages targeting specific employees based on their online activity, thereby increasing the probability of success (Kaloudi et al., 2020). The ISO/IEC 27001 and NIST models can help counter the traditional threats, but are not quite robust in the context of integrating AI-based defenses into the current security architectures (Malatji, 2023).

For MFIs to identify threats in real-time and detect suspicious activities, AI-based solutions like machine-learning-based threat detection and behavioral analytics are essential. The integration of AI-based cybersecurity technologies can help MFIs move from a reactive defense to a proactive cybersecurity approach (Ali et al., 2023). These technologies can stop phishing attacks by automatically identifying suspicious activity and investigating staff behavior patterns before a breach occurs. For example, behavioral analytics can identify an anomaly in normative user behavior, such as unusual login times or unusual devices, and provide early alerts of a possible attack (Javaheri et al., 2023).

3.3.1 Collaborative Cybersecurity Models

Collaborative cybersecurity models are an economically feasible option for dealing with cybersecurity issues, especially in light of micro-finance institutions' (MFIs) limited budget. MFIs can also engage expert vendors to perform key cybersecurity tasks, such as vulnerability management, incident response, and threat detection, through cybersecurity as a service (CaaS) (Morris et al., 2023). Consequently, this strategy makes MFIs achieve real-time monitoring and better threat intelligence without having to spend significant amounts of money on costly cybersecurity equipment (Gupta et al., 2021).

For instance, a group of MFIs in East Africa can build a common cybersecurity infrastructure, provide centralized threat detection services, and thereby distribute the burden of maintaining separate systems among individual institutions (Ali et al., 2023). Joint ventures with governmental organizations or non-governmental organizations (NGOs) may provide the technical support, funding, and training in order to enhance the cybersecurity defenses of MFIs. Becoming members of collaborative frameworks allows MFIs to overcome extensive financial and expertise barriers to implementing resilient cybersecurity measures (Xu et al., 2020).

3.3.2 Regulatory Compliance Challenges for MFIs

Microfinance institutions (MFIs) face serious challenges in complying with local data protection laws and regulations, such as the General Data Protection Regulation (GDPR). Sensitive information acquisition, storage, and transmission require strong controls and compliance with these rules, which usually call for a significant investment in cybersecurity infrastructure (Shafiq et al., 2020). As an example, the GDPR imposes incident reporting, privacy-by-design standards, and data encryption, which might be beyond the financial and technological capabilities of smaller MFIs (Gal et al., 2020).

In addition, MFIs face challenges in conducting compliance audits and maintaining ongoing adherence to changing regulatory requirements. Failure to do so can attract heavy fines and legal repercussions, hence straining the already tight budget (Xu et al., 2020). Turning to government-subsidized programs or partnering with compliance-as-a-service (CaaS) agencies can be a reasonable approach, providing affordable solutions to ensure regulatory compliance without disrupting operations (Morris et al., 2023).

3.3.3 Measuring the Success of Cybersecurity Strategies

To evaluate the effectiveness of cybersecurity measures, microfinance institutions (MFIs) should set clear key performance indicators (KPIs) to track the success of their strategies. The reduction in phishing attacks, the rate of compliance improvements, and the speed of incident response are KPIs that can give objective information on the effectiveness of cybersecurity frameworks (Bardhan et al., 2021). MFIs can also monitor operational availability after a cyberattack or data breaches to assess the speed and resiliency of their recovery (Sicard, 2019).

For example, MFIs that use AI-based phishing detection mechanisms can assess the effectiveness of these tools by tracking false-positive and detection rates. Similarly, compliance audits can be used to measure how well organizations adhere to the GDPR and other regulations, thereby helping avoid financial penalties and ensuring compliance (Shafiq et al., 2020). With clearly defined success metrics, MFIs can periodically evaluate and improve their cybersecurity posture as new threats arise.

3.3.4 Global Cybersecurity and AI Governance Frameworks

With the increasing complexity of phishing threats, especially AI-generated attacks, there is a growing need for a comprehensive governance framework that goes beyond traditional cybersecurity controls to address considerations such as algorithmic integrity, explainability, and accountability. Traditional standards, such as the NIST Cybersecurity Framework (2020) and ISO/IEC 27001 (2022), are still useful in institutional risk management; however, they are rule-oriented and do not offer any mechanisms to regulate the design, implementation, and ethical deployment of artificial intelligence in the processes of cybersecurity (Veale & Borgesius, 2021).

These restrictions are particularly important for complex neural networks, such as Gated Recurrent Units (GRUs) used in phishing mitigation, because they rely heavily on automation and offer limited transparency.

In terms of responsive regulation (Ayres & Braithwaite, 2023), regulatory frameworks should strike a balance between innovation and compliance by strengthening their control in light of the risks posed by emerging technologies. The emerging EU Artificial Intelligence Act is an example of this method, as it identifies phishing detection in the finance sector as a high-risk AI application, which is why it is subject to compulsory documentation, bias audits, and human control mechanisms (European Commission, 2021). The Act presents a normative prototype of risk-tiered regulation, under which technologies capable of causing severe harm to society should be subject to more intensive scrutiny. Likewise, the Principles on the Ethical Use of AI and Data Analytics of the Monetary Authority of Singapore (MAS) provide a domain-specific, principles-based framework to ensure that financial AI systems are grounded in fairness, ethics, accountability, and transparency (MAS, 2021).

The governance frameworks discussed in this section were identified through the systematic literature review undertaken for this study. The detailed review protocol, including the literature sources, screening process, and study selection procedures, is presented in Section 3.5. Table 2.1 compares the major international cybersecurity and AI governance frameworks identified through the review and examines their relevance to phishing mitigation and AI-enabled cybersecurity. The comparison indicates that widely adopted frameworks, including ISO/IEC 27001 and the NIST Cybersecurity Framework, provide limited guidance on important AI governance considerations such as algorithmic fairness, model interpretability, and auditability. These aspects have become

increasingly important as artificial intelligence is deployed to support cybersecurity decision-making in high-risk sectors such as financial services.

Table 3.1: Cooperative Overview of Global Cybersecurity and AI Governance Frameworks

Framework	**Jurisdiction**	**AI Relevance**	**Phishing-Specific Provisions**
NIST CSF (2020).	United States	None	Limited to identity/access controls
ISO/IEC 27001 (2022)	Global	None	Procedural information security only
GDPR (2021 updates)	European Union	Indirect references	Data privacy and consent mechanisms
EU Artificial Intelligence Act (2021).	European Union	Risk-tiered AI regulation	High-risk classification for phishing detection
MAS FEAT Principles (2021)	Singapore	AI-specific ethics	Designed for financial services
NITI AI Strategy (2021).	India	Broad ethical framing	No phishing-specific guidance
AI Ethics Framework (2020).	Australia	General principles	Lacks enforcement tools

AI Security Guidelines (2023 Draft)	China	Centralized risk model	Limited detail

Source: Author (2025).

Although these frameworks have strong normative content, their enforcement mechanisms differ significantly. The EU AI Act is a draft proposal that can be effected or put on hold before implementation. Likewise, the MAS Principles are not binding and have to be voluntarily enforced, which casts doubt on the regulatory authority. In less-developed countries, the institutionalization of AI governance is often hindered by limited regulatory capacity, technical expertise, and political will, which, according to the institutional theory of global norms, decisively affect the choice of global norms (DiMaggio & Powell, 1983).

Kenyan MFIs operate in highly resource-limited environments and may not be in a position to fully meet international guidelines. However, a hybrid strategy that merges the strengths of the EU AI Act's risk-based approach and the MAS FEAT Principles' sector specificity may offer a blueprint for contextual implementation. The model should be adapted to local technological capabilities, legal frameworks, and regulatory maturity, so that AI-assisted cybersecurity can be practiced in a sustainable and responsible way without imposing unreasonable compliance requirements.

3.3.5 Kenya’s Cybersecurity Legal Framework and AI Governance Gaps

Kenya has made significant strides in developing a legal framework for data and cybersecurity management. The primary legal frameworks aimed at protecting data integrity, punishing cybercrime, and facilitating the institution's readiness are the Data Protection Act (2019), the

Computer Misuse and Cybercrimes Act (2018), and the National Cybersecurity Strategy (2014). However, these instruments are substantively ineffective at addressing the unique challenges of AI-based cybersecurity governance.

From a risk-based regulatory perspective (Gonçalves, 2020), AI-based phishing mitigation systems, especially those based on gated recurrent units (GRUs), should be deemed high risk due to their impact on access to financial services and their susceptibility to bias and misclassification. Nevertheless, Kenyan law does not address the circumstances under which such models should be audited, certified, or clarified. The Computer Misuse and Cybercrimes Act outlaws phishing attacks but leaves no regulatory oversight of the tools used to prevent them; Johnson (2021) writes that this is an implementation asymmetry: banning particular results without setting any rules for how to prevent them.

Table 2.2 shows the discrepancy between the measurement of significant cybersecurity laws in Kenya and AI preparedness. There is no provision for algorithmic governance in any statute, which shows that the legal architecture for the deployment of AI is incomplete.

Table 3.2 Kenya's Cybersecurity Legal Frameworks and Gaps for AI Governance

Law/Policy	**Primary Strength**	**Gap in AI Governance**
Data Protection Act (2019)	Privacy, consent, data minimization	No clauses on automated decision-making or model auditing
CMCA (2018).	Phishing criminalization	No regulation for AI-based detection tools

National Cybersecurity Strategy (2014).	Inter-agency coordination	Pre-AI framework; lacks ML policy integration
KICA (1998)	ICT regulation	Outdated and lacks an AI reference

Source: Author (2025).

Responsive regulation-based theoretical viewpoints suggest that in the case of legal gaps, the government can step in by creating regulatory sandboxes, i.e., temporary supervised settings where artificial intelligence (AI) systems can be tested before full-scale implementation. One in particular is the sandbox model that was adopted by the Monetary Authority of Singapore, which contributed to fintech innovation without compromising oversight (MAS, 2021). Introducing and piloting similar sandbox platforms by Kenyan regulatory authorities (e.g., the Communications Authority (CA) and the National Kenya Computer Incident Response Team - Coordination Centre (KE-CIRT/CC), in collaboration with the Central Bank of Kenya (CBK), for AI-based cybersecurity tools in the financial sector could be considered.

In addition, Kenya can align its domestic regulatory policy with regional standards, such as the Malabo Convention (African Union, 2014), and with national regulations similar to the South Africa Cybercrimes Act (2021) to address issues such as algorithmic accountability, data localization, and AI ethics. The need to develop policy coherence at the regional level is timely and crucial, as the African Union Digital Transformation Strategy (2020-2030) underscores the importance of AI preparedness.

Despite the limited resources, the policy reforms can be implemented gradually. Preliminary steps could include requiring algorithmic transparency reporting under the Data Protection Act, creating an AI ethics advisory panel within the Kenya Bureau of Standards (KEBS), and establishing AI governance contexts between public and private partners, including academics and civil society organizations. Without these measures, Micro-Finance Institutions risk implementing GRU-based phishing detection models without a regulatory framework, exposing them to legal, ethical, and reputational risks.

In conclusion, the existing legal and policy framework in Kenya cannot be used to govern AI-based cybersecurity measures. However, it is necessary to have a consistent, risk-sensitive, and practice-based regulatory response to ensure that AI delivers benefits without undermining digital trust in the financial sector.

3.4 Contemporary AI-driven models for detecting phishing attacks

Phishing continues to be a steady threat to digital financial systems, particularly in the micro-finance institutions (MFIs), where sensitive user and transaction data are at risk. Traditional detection methods, such as rule-based filters and signature-based algorithms, could not identify evolving phishing patterns, such as spear-phishing and zero-day exploits (Olabiyi, 2024). However, deep-learning methods exhibit dynamic adaptability by learning from large, heterogeneous datasets to uncover subtle indicators of malicious intent. It has been empirically shown that neural-network-based models are better than traditional methods in identifying anomalies in e-mails, URLs, and user-behavior patterns (Xu et al., 2020). However, the use of AI-based applications in MFIs is still hindered by limited computational resources, a shortage of skilled technical personnel, and the need to comply with strict data protection laws (Somesha et

al., 2020). Regardless of these challenges, AI-based phishing-detection systems promise to be an ultimate breakthrough towards proactive cybersecurity resilience in financial ecosystems.

3.4.1 Deep Neural Network Models in Cybersecurity

Advancements in deep learning architectures have significantly improved phishing detection, with different architectures showcasing their strengths. Convolutional Neural Networks (CNNs), initially designed for image detection, have proven effective in analyzing the content structures of e-mails and websites to detect phishing attempts (Aljofey et al., 2020). CNNs are good at identifying patterns; they perform structural analysis of e-mails and websites to detect phishing attempts. Nevertheless, temporal dependencies, which often play a critical role in phishing detection scenarios, are difficult for CNNs to capture (to include citation here).

This weakness can be addressed by Recurrent Neural Networks (RNNs), which are especially effective for sequential data. RNNs capture temporal dependencies in user actions or email behaviors and thus can provide significant value for phishing detection. However, the vanishing gradient problem hinders RNNs' ability to maintain long-term dependencies, diminishing their applicability to complex, long-term interactions (Noh, 2021).

To overcome these shortcomings, Long Short-term Memory (LSTM) and bidirectional Long Short-Term Memory (Bi-LSTMs) were developed. LSTMs use various gates, including input, forget, and output gates, to control the flow of information, helping manage long-term dependencies within the network. Since the e-mail interaction patterns can change as the number of interactions increases, LSTMs are effective in identifying phishing (Alshingiti et al., 2023). By

capturing bi-directional relationships, Bi-LSTMs, the sequence analyzers working in both forward and backward directions, are even more effective, but they require more computational resources (Schuster & Paliwal, 2020).

GRUs are a simple variant of LSTM that is used to reduce the computational complexity while maintaining the vital long-term dependencies. GRUs ensure that unnecessary information is eliminated by modulating data retention using update and reset gates, thus improving computational efficiency. GRUs are therefore best suited in settings with limited resources like MFIs (Roy et al., 2022). GRUs can maintain crucial temporal dependencies, and their ability to accelerate processing makes them highly suitable for real-time phishing detection.

3.4.1.1 GRU Architectural Overview and Computational Advantages

Gated Recurrent Units (GRUs) simplify the recurrent architecture by using only two gates—the update and reset gates—compared to the three gates in LSTMs. The update gate determines how much past information should be carried forward, while the reset gate controls how new input is combined with the previous memory. This streamlined mechanism reduces parameter count and accelerates training without significantly compromising accuracy (Roy et al., 2022).

Mathematically, the GRU's operation can be expressed as follows:

$$z_t = \sigma\,(W_z\, x_t + U_z\, h_{t-1}) \qquad \textit{Equation 3.1}$$

$$r_t = \sigma\,(W_r\, x_t + U_r\, h_{t-1}) \qquad \textit{Equation 3.2}$$

$$h_t = (1 - z_t)\, h_{t-1} + z_t \tanh\,(W_h\, x_t + U_h\,(r_t\, x\ \ h_{t-1})) \qquad \textit{Equation 3.3}$$

Here, z_t and r_t represent the update and reset gates, respectively, x_t is the input at time t, and h_t is the current hidden state. The simplicity of these equations highlights why GRUs are computationally lighter and more suitable for real-time phishing detection in MFIs.

Unlike LSTMs, which explicitly separate memory and hidden states, GRUs merge them, reducing the memory footprint while maintaining strong performance in sequential analysis (Alshingiti et al., 2023). This design balance makes GRUs particularly valuable in domains where both speed and interpretability are critical.

3.4.1.2 Comparative Analysis: RNNs, LSTMs, GRUs, and Transformers in Phishing Detection

The evolution of sequential learning—from basic RNNs to advanced Transformer models—reflects a continual effort to enhance contextual understanding and computational efficiency in phishing detection. Traditional RNNs introduced the notion of sequential dependencies, but their susceptibility to vanishing gradients limited their ability to learn long-term dependencies (Hochreiter, 2020). LSTMs addressed this by introducing gated control, improving memory stability, and enabling better detection of contextual phishing cues across email threads and user behavior. GRUs streamlined this process by merging the input and forget gates, reducing training time and computational overhead (Roy et al., 2022).

More recently, Transformer architectures have redefined sequence modeling using self-attention mechanisms that process entire input sequences simultaneously rather than step by step (Tay et al., 2021). This design enables a more holistic understanding of contextual relationships in phishing content, including the detection of impersonation and semantic inconsistencies (Li et al., 2024). However, Transformers demand substantial hardware resources and are less suited to institutions

with limited infrastructure, such as MFIs. To address this, hybrid models that integrate GRUs with lightweight attention mechanisms have emerged as a promising middle ground, offering improved contextual sensitivity with manageable computational costs (Chen et al., 2023).

Overall, RNNs laid the foundation for sequential modeling; LSTMs improved memory retention; GRUs optimized computational efficiency; and Transformers revolutionized contextual comprehension. The choice of architecture depends on balancing accuracy, efficiency, and resource availability within cybersecurity frameworks.

Table 3.3 Comparative Characteristics of RNN, LSTM, GRU, and Transformer Architectures in Phishing Detection Applications

Parameter	**RNN (Recurrent Neura l Network)**	**LSTM (Long Short-Term Memory)**	**GRU (Gated Recurrent Unit)**	**Transformers**
Architecture	Simple recurrent structure with feedback loops.	Incorporates memory cells with input, forget, and output gates.	Combines input and forget gates into an update gate; fewer parameters.	Utilizes a self-attention mechanism and feed-forward layers without recurrence.
Handling Long Sequences	Struggles to maintain long-term	Excels in capturing long-term	Better than RNNs but slightly less	Effectively manages long sequences using

	dependencies due to vanishing gradients.	dependencies through gated memory control.	effective than LSTMs at handling long-term dependencies.	self-attention mechanisms that capture global context.
Training Time	Fast but less accurate on complex data patterns.	Slower due to multiple gates and memory operations.	Faster than LSTMs but slower than RNNs due to moderate gate complexity.	Requires heavy computation but supports parallelized training.
Memory Usage	Low memory requirements and light computation.	Higher memory consumption due to complex architecture.	Lower memory usage compared to LSTMs but higher than RNNs.	High memory usage because of multi-head attention and large feed-forward layers.
Parameter Count	Fewer parameters overall.	More parameters than RNNs due to three gates and memory cells.	Fewer parameters than LSTMs owing to a simplified two-gate structure.	A very large number of parameters in the multi-head attention and

				positional encoding layers.
Ease of Training	Prone to vanishing gradients, making it harder to train on long sequences.	Easier to train on long sequences due to controlled gradient propagation.	Simpler and easier to train than both RNNs and LSTMs.	Requires significant computational power, GPUs, and large-scale data.
Use Cases	Suitable for simple sequence modeling such as stock forecasting or basic event detection.	Ideal for time-series forecasting, text generation, and tasks requiring long-term dependencies.	Preferred in low-resource environments for efficient real-time phishing detection and spam classification.	Best suited for advanced natural language processing and cybersecurity applications such as phishing detection, threat intelligence, and behavioral anomaly analysis.
Parallelism	Limited due to sequential	Similar limitations as RNNs;	Same limitations as RNNs and LSTMs, but	High parallelism is enabled by attention-based

	computation constraints.	sequential data flow restricts parallelization.	slightly faster convergence.	and non-sequential design.
Performance on Long Sequences	Poor performance on long sequences.	Strong performance and stable memory over long sequences.	Moderate to good performance on long sequences.	Excellent performance on long sequences due to attention-driven context modeling

As shown in Table 2.3, each model type offers trade-offs between accuracy, interpretability, and computational requirements. Transformers achieve superior contextual performance, but GRUs provide a pragmatic compromise for real-world microfinance cybersecurity implementations.

3.4.2 Evaluation of Deep Neural Network Models

GRUs have shown significant benefits over conventional recurrent neural networks (RNNs) and even Long Short-Term Memory (LSTM) architectures in phishing detection. Although LSTMs are effective at capturing long-term dependencies, they require longer training times and greater computational resources. GRUs, on the other hand, can handle long-term dependencies because they require fewer parameters, leading to lower computational complexity. GRUs, in turn, are particularly effective for real-time phishing detection, where rapid adaptation is essential (Roy et al., 2022).

While Bidirectional LSTMs (Bi-LSTMs) enable bidirectional sequence processing, they are not practical in settings such as microfinance institutions (MFIs), where resources may be constrained by high computational requirements. GRUs are thus the best choice for phishing detection in MFIs, as they offer a better trade-off between computational efficiency and performance.

3.4.3 Use of GRUs in Phishing Detection

The GRU model's ability to process sequential information makes it particularly effective for phishing detection. GRUs are more effective in detecting phishing and hidden anomalies compared to traditional detection systems by analyzing patterns in email content, sender behavior, and the interaction with embedded links and URLs (KS, 2024). To combat ever-changing phishing techniques, empirical research has shown that GRUs are highly accurate at detection by capturing long-term dependencies in user behavior and email exchanges (Huang et al., 2020). GRUs are the best option since they have low computational overhead when an organization's technological resources, such as those of microfinance institutions (MFIs), are limited.

3.4.4 Challenges and Limitations of GRUs in Cybersecurity

Despite these benefits, GRUs have several limitations in phishing detection and broader cybersecurity applications. It is hard to build extensive datasets that cover all possible attack vectors, as phishing methods evolve rapidly (Alkhalil et al., 2021). In addition, despite GRUs' ability to identify advanced phishing patterns, the false-positive rate remains a major issue, as valid emails can be misclassified as suspicious, causing operational disruptions (Cahuantzi et al., 2023). Although GRUs are more computationally efficient than LSTMs, they still require substantial

processing power, which can be an issue for smaller facilities with limited technological capacity (Cahuantzi et al., 2023).

Another critical concern is how the model is prone to adversarial attacks, in which attackers alter inputs to cause misclassification. This highlights the need to conduct further research to make GRU-based phishing detection systems more resilient (Do et al., 2022). To address the unreliability of GRUs in practical applications, it was necessary to develop countermeasures against such adversarial attacks.

3.4.5 Explainable Artificial Intelligence (XAI) in AI-driven Phishing Detection

The increasing adoption of artificial intelligence (AI) and deep learning techniques in cybersecurity has significantly enhanced the ability of organizations to detect sophisticated phishing attacks. Deep learning models, including recurrent neural networks (RNNs) and gated recurrent units (GRUs), have demonstrated strong capabilities in identifying complex patterns within email content, user behavior, and malicious communication characteristics. However, despite their high predictive performance, many deep learning models operate as black-box systems, where the reasoning behind their predictions is difficult for human users to understand. This lack of transparency presents a significant challenge, particularly in cybersecurity environments where decisions regarding threat classification, incident response, and risk mitigation require confidence, accountability, and human validation.

Explainable Artificial Intelligence (XAI) has emerged as an approach for improving the transparency and interpretability of AI-driven systems by providing understandable explanations of model predictions. Unlike conventional AI models that only provide an output, such as

classifying an email as phishing or legitimate, XAI techniques attempt to reveal the factors contributing to the decision. In phishing detection, explainability can assist users and cybersecurity professionals in understanding why a particular message was identified as malicious by highlighting influential features such as suspicious URLs, abnormal sender characteristics, unusual linguistic patterns, misleading content, or deviations from normal communication behaviour.

The importance of XAI in cybersecurity extends beyond improving model transparency. In operational environments, security analysts often require evidence to validate automated decisions before initiating defensive actions. A highly accurate AI model without interpretability may reduce user confidence, increase resistance to adoption, and limit its practical usefulness. This challenge is particularly relevant in digital microfinance institutions, where cybersecurity decisions affect financial transactions, customer information, and institutional trust. Therefore, explainability can support more informed decision-making by transforming complex AI outputs into actionable security insights that can be understood by both technical and non-technical stakeholders.

Several XAI approaches have been proposed to improve the interpretability of machine learning and deep learning models. Model-agnostic techniques, such as Local Interpretable Model-Agnostic Explanations (LIME) and SHapley Additive Explanations (SHAP), provide explanations by estimating the contribution of individual input features to a model's prediction. These approaches can be applied to different AI architectures without requiring modifications to the underlying model. In addition, model-specific approaches, such as attention mechanisms in neural networks, allow researchers to identify important sections of input sequences that influence classification decisions. For phishing detection systems, these techniques can provide insights into the characteristics that cause an email or message to be classified as malicious.

Although XAI improves transparency and trust, its application in cybersecurity also presents challenges. Excessive disclosure of model behavior may provide attackers with information that could be exploited to evade detection mechanisms. Furthermore, explanations generated by complex AI models may themselves be difficult for non-technical users to interpret, reducing their practical value. Therefore, effective XAI implementation requires a balance between providing meaningful explanations and maintaining security against adversarial manipulation.

From a microfinance cybersecurity perspective, XAI provides an important link between advanced AI capabilities and human-centered security operations. Digital microfinance institutions often operate within environments where cybersecurity expertise, resources, and technical capacity may be limited. Providing understandable explanations for AI-generated phishing alerts can improve users' trust in automated security systems and support more effective incident response. This aligns with the socio-technical view of cybersecurity, where technological solutions must be integrated with human decision-making processes and organizational practices.

Furthermore, integrating XAI into adaptive AI-based phishing detection frameworks enhances accountability and continuous improvement. As AI models learn from emerging phishing techniques, explainability mechanisms can assist security teams in evaluating model behavior, identifying weaknesses, and refining detection strategies. Therefore, XAI complements adaptive neural network approaches by ensuring that improvements in detection performance are accompanied by transparency and operational reliability.

Consequently, this study considers Explainable Artificial Intelligence as an important supporting capability within AI-driven phishing detection frameworks. While the GRU-based model provides the capability to learn and classify complex phishing patterns, XAI enhances the interpretability

of these decisions by providing meaningful explanations to users and cybersecurity practitioners. The integration of explainability addresses a critical limitation of deep learning-based cybersecurity solutions and contributes towards developing a trustworthy, adaptive, and practically deployable phishing resilience framework for digital microfinance institutions.

3.4.6 Integration of GRUs into MFI Cybersecurity Frameworks

To enable real-time analysis, deploying GRUs would require major improvements in data management, hardware, and software.

The integration of Gated Recurrent Unit (GRU) models into modern cybersecurity systems creates interoperability and resource allocation issues for microfinance institutions (MFIs). The signature-based detection methods, which are still used by many MFIs, are incompatible with AI-based approaches such as GRUs (Rizvi et al., 2022). Deploying GRU would require significant improvements in data management, hardware, and software infrastructure to enable real-time analysis.

A lack of machine learning and AI administration expertise also contributes to barriers to adoption. Most MFIs are now facing high financial costs because they may be forced to outsource cybersecurity services or invest heavily in training (Alam, 2024). Furthermore, because GRU models must comply with strict data protection laws and remain accountable and transparent, integrating them into current regulatory and compliance frameworks is also challenging (Khaleefah et al., 2024).

Future studies should concentrate on strengthening GRUs' limitations on these challenges and researching methods for smoother integration into MFI cybersecurity systems.

In conclusion, despite GRUs' promise of significant progress in phishing detection for microfinance institutions, their integration into existing cybersecurity systems is fraught with challenges. GRUs can be applied especially to resource-limited settings due to the capability to balance computational efficiency and performance; however, the problem of false positives, adversarial attacks, and the need to have reliable datasets need to be resolved. Future studies ought to consider addressing the shortcomings of GRUs with regard to these issues and devising ways for smoother integration into MFI cybersecurity systems.

3.4.7 Advancing GRUs Integration Through Hybrid and Collaborative Models

The integration of Gated Recurrent Units (GRUs) into current cybersecurity frameworks is one of the opportunities the microfinance institutions (MFIs) have to increase phishing detection and cybersecurity resilience in general. However, this process has a number of challenges with regard to allocation of resources, interoperability, and regulatory standard compliance. GRUs, which are a type of Recurrent Neural Networks (RNNs), are best suited to phishing detection because of their computational efficiency and their ability to detect long-term dependencies in sequential data. Such features make GRUs useful in real-time phishing detection in resource-constrained environments; nevertheless, to incorporate them into the existing systems of MFIs necessitates careful preparation and adjustment.

Interoperability with pre-existing legal systems is one of the main barriers. A significant number of MFIs still use old-fashioned signature-based detection methods that cannot counter modern, AI-driven attacks like GRUs. GRU-based models would necessitate massive infrastructure investment, including upgrades to hardware, software, and the data management system to enable real-time analytics (Rizvi et al., 2022). Such improvements can include the use of high-

performance computing solutions or the implementation of cloud-based AI technologies to handle massive amounts of data within a short period of time.

Another serious impediment is the lack of in-house knowledge in AI and machine learning (ML) administration. Most MFIs do not have the technical personnel required to service and install sophisticated AI systems like GRUs. As a result, they are often forced to outsource cybersecurity services, which increases the cost of integration (Alam, 2024). MFIs can gain the necessary technical expertise required for successful GRU deployment through the help of capacity-building programs or collaboration with cybersecurity service providers or Fintech start-ups. Additionally, MFIs can access advanced AI-driven phishing detection using cybersecurity-as-a-service (CaaS) models by outsourcing key cybersecurity tasks, which can be a cost-effective option (Morris et al., 2023).

There are other regulatory and compliance challenges associated with the incorporation of GRUs into existing frameworks. GRUs must operate under strict data-protection laws that should be observed by MFIs. Although this operational transparency should be maintained, the integration process should ensure that the GRUs comply with data privacy requirements, such as those provided by the GDPR or ISO/IEC 27001. During phishing detection, sensitive data can be securely handled by the use of privacy-preserving AI systems, such as federated learning and differential privacy (Camacho, 2024). Moreover, ensuring compliance and resolving issues about false positives or system failures is largely determined by the transparency of the AI models, especially in the context of decision-making (Khaleefah et al., 2024).

Despite these challenges, there are distinct advantages to incorporating GRUs into MFI cybersecurity frameworks to curb phishing attacks. GRUs accurately identify real-time threats by

detecting trends in user behavior, email content, and URL interactions, using their consolidated gating mechanisms. Additionally, they are able to dynamically adjust to the new threats and update detection algorithms in response to the new phishing techniques (Roy et al., 2022). GRUs are more computationally efficient, making them better alternatives to more resource-intensive models such as Bi-LSTMs or Multilayer Perceptron (MLP) in settings with limited resources, such as MFIs (Ali et al., 2023).

However, several financial and technical obstacles must be overcome before GRUs can be successfully integrated into MFI cybersecurity frameworks. A phased integration strategy, in which MFIs continue to employ traditional detection methods while progressively implementing GRU-based models, might be efficient. This reduces the immediate financial impact and enables a more seamless transition. Furthermore, MFIs may be able to exchange real-time threat intelligence and pool resources through collaborative cybersecurity frameworks, which would make the incorporation of GRUs more practical and economical for smaller institutions (Shafiq et al., 2020).

Another major obstacle to GRU integration is cost, since implementing these models necessitates ongoing system maintenance in addition to hardware and software investments. By removing the requirement for extensive on-site infrastructure, cloud-based AI solutions or cybersecurity-as-a-service models can allow MFIs access to AI-driven phishing detection and reduce these expenses. This reduces initial capital costs while offering flexibility and scalability to meet evolving cybersecurity needs (Morris et al., 2023).

In conclusion, there is a lot of potential for boosting overall security and phishing detection by integrating GRUs into MFI cybersecurity frameworks. Even though the process includes

challenges such as incompatibility with legacy systems, a lack of technical expertise, and regulatory obstacles, these can be overcome with cloud-based solutions, collaborative frameworks, and staged deployment plans. Adopting AI-driven solutions like GRUs was essential to guaranteeing MFIs' long-term cybersecurity resilience as phishing attacks evolve.

3.5 Case Studies of Phishing Attacks in Microfinance

The rising reliance on digital financial services within microfinance institutions is a growing concern for phishing attacks. As such, institutions with limited resources are particularly vulnerable to sophisticated phishing schemes that often rely on technical tactics and exploit human vulnerabilities. Phishers attack employees as well as customers, which can lead to serious financial losses, data breaches, and reputation damage. This section examines some of the largest phishing attacks in Kenya, India, and Nigeria. We'll see how these attacks happened, the damage they caused, and how the companies and institutions tried to stop them. These cases were also compared to similar attacks in other developing countries to understand the bigger picture and find ways to protect against them.

3.5.1 Analysis of High-Profile Phishing Cases in Microfinance Institutions

Case Study 1: Kenya (2020)

In 2020, a major microfinance institution in Kenya fell victim to a phishing attack that leveraged social engineering via email. Cybercriminals posed as internal staff members and sent fraudulent emails to employees and clients, directing them to click on malicious payment alert links. When the links were clicked, the victims were taken to a page designed to steal their credentials. The attack caused significant financial losses, as many transactions were executed using compromised

client accounts. Even with efforts to contain the breach, quick detection was not possible due to the lack of automated systems to mitigate online phishing attacks, leading to significant financial and reputational losses for the MFI. This incident is consistent with broader patterns in Kenya, where phishing attacks are increasingly targeting digital financial platforms and exploiting weaknesses in the cybersecurity infrastructure (Ibrahimnur, 2023).

Case Study 2: India (2021)

In 2021, one of the largest micro-financial institutions in India was hit by a hybrid email phishing attack that leveraged social engineering tactics. The attackers, posing as technical support, sent fraudulent emails to employees, asking them to share their login credentials under the pretense of an IT-related issue. This resulted in a serious data breach in which sensitive client data, including personal identification and financial details, was compromised. The regulatory response in India was swift, with the government collaborating to enhance cybersecurity measures and raise awareness across the sector. However, the breach severely impacted client trust and required significant recovery efforts from the institution. Similar incidents have been documented across India's digital microfinance sector, with underprivileged groups increasingly targeted by data theft and phishing (Afzal et al., 2024).

Case Study 3: Nigeria (2022)

In 2022, a phishing attack using a URL-based technique targeted a large, city-based microfinance institution in Nigeria. The attackers designed a fraudulent website that closely resembled the MFI's official portal to lure clients into entering their login credentials. This served its purpose in compromising the login credentials, which were later used to execute unauthorized withdrawals

from clients' accounts. Even though the institution attempted to mitigate the damage, the attack still caused significant monetary losses and reputational damage in the long run. In Nigeria, other stakeholders, such as regulatory bodies, stepped in to address the situation by implementing more stringent cybersecurity measures in the microfinance sector (Nwankwo et al., 2023). This is consistent with broader research indicating that Nigerian MFIs are more vulnerable to phishing, social engineering, and bogus websites due to rapid digitization and lax compliance (Daudu et al., 2025).

3.5.2 Analysis of the phishing attack Techniques in the three cases

Phishing attacks across all three cases utilized social engineering as a primary attack vector. In Kenya and India, attackers exploited human vulnerabilities by impersonating trusted entities within the MFI, such as IT support or internal management. These phishing emails were designed to create a sense of urgency, prompting victims to respond quickly and without verifying the legitimacy of the requests. In contrast, the Nigerian case focused on a more direct attack on clients through URL manipulation. Clients were redirected to a fake website that closely resembled the MFI's official portal, where they unknowingly shared their login credentials (Alkhalil et al., 2021).

The attacks revealed technical vulnerabilities within these MFIs, including the lack of multi-factor authentication (MFA), weak email security protocols, and insufficient use of phishing detection systems. In all cases, machine learning (ML)- based phishing detection mechanisms were either underutilized or absent. Advanced ML models such as Gated Recurrent Units (GRUs) and Convolutional Neural Networks (CNNs), which have been shown to effectively detect phishing attempts by analyzing anomalies in email content and URLs, could have significantly reduced the success of these attacks (Roy et al., 2022). The absence of such systems highlights the need for

improved technical defenses to complement human-centered security measures. The table below presents an analysis of the case studies, highlighting how each country handled the various phishing attacks.

Table 3.4 Summary of Case Studies

Criteria	**Case Study 1**	**Case Study 2**	**Case Study 3**
Location	Kenya	India	Nigeria
Year of Incidence	2020	2021	2022
Type of Attack	Phishing (email-based, internal impersonation)	Phishing (email, social engineering)	Phishing (URL-based, fake website)
Type of MFI	Large national MFI	Leading regional MFI	Major urban MFI
Techniques Used to Attack	Email phishing, URL redirection	Email phishing, social engineering	URL manipulation, phishing website
Impact – Financial	Significant financial losses, unauthorized withdrawals	Data theft, exposure of client financial details	Unauthorized withdrawals, loss of client trust
ML Techniques Used to Detect	None (manual detection)	None (manual detection)	Basic URL filtering
ML Technique Defense Used	N/A	N/A	Basic URL filtering

Success Rate of Defence Technique	Low (delayed detection, manual processes)	Low (no ML techniques)	Low (basic filtering was bypassed)
Response (Govt, MFI, Regulatory)	Regulatory intervention, cybersecurity upgrade	Swift regulatory response, enhanced security policies	Regulatory action, client compensation, and efforts

(Source: *Author's own compilation)*

3.5.3 Impact and Defense Mechanisms in all three MFIs

The financial and reputational impacts of these phishing attacks were severe across all cases. In Kenya, the unauthorized withdrawals and delayed detection led to major financial losses, further aggravated by the institution's slow response. In India, the breach of client data posed long-term risks to the privacy and security of thousands of clients, while in Nigeria, the attack caused widespread client dissatisfaction and loss of trust in the MFI's ability to safeguard their assets. Despite efforts by these institutions to recover from the attacks, including client compensation and internal cybersecurity upgrades, the damage was significant (Nadeem et al., 2023).

In terms of defense mechanisms, all three MFIs relied on manual detection processes, which proved ineffective against sophisticated phishing schemes. Basic URL filtering and email security measures failed to detect the attackers' advanced social engineering techniques. While none of the institutions had fully implemented ML-based phishing defenses, the Nigerian MFI had a basic URL filtering system that was ultimately bypassed. More advanced ML techniques, such as behavioral analysis through GRUs or CNNs, could have identified the subtle anomalies in email

content or login activity, thereby improving detection rates and reducing the impact of the attacks (Gandotra et al., 2021).

3.5.4 Lessons Learned from the Case Studies

Several critical lessons emerge from these case studies. First, human vulnerabilities remain the primary target for phishing attacks. Employees and clients must be trained to recognize phishing attempts, and regular cybersecurity awareness programs should be implemented to reduce susceptibility to social engineering attacks (Hillman et al., 2023). Second, the absence of machine learning-based defenses in phishing detection was a significant oversight in all cases. The adoption of ML techniques, such as GRUs and CNNs, can significantly improve phishing detection by identifying subtle patterns in email and URL structures that are difficult for traditional systems to detect (Roy et al., 2022). Third, multi-factor authentication (MFA) is essential for mitigating phishing risks. In all three cases, the absence of MFA allowed attackers to easily compromise accounts once they obtained credentials.

Furthermore, the role of regulatory bodies in responding to phishing attacks is crucial. Governments and financial authorities must ensure that MFIs comply with minimum cybersecurity standards, including the use of MFA, regular security audits, and the implementation of ML-based detection systems (Dawson et al., 2022). Regulatory measures in Kenya and Nigeria, while eventually enacted, were delayed, exacerbating the impact of the attacks. Swift regulatory intervention, as seen in the Indian case, can help limit the damage and restore client confidence.

Phishing attacks on MFIs, as evidenced by the case studies in Kenya, India, and Nigeria, highlight the vulnerabilities that these institutions face in a rapidly digitalizing financial landscape. Attackers

exploit both human error and technical weaknesses, leading to significant financial losses and reputational damage. The absence of machine learning-based phishing detection and multi-factor authentication in these cases underscores the need for MFIs to adopt more advanced cybersecurity measures. Incorporating the best global practices, particularly the use of ML-based anomaly detection systems and behavioral analysis, can strengthen phishing defenses in MFIs across developing regions. In addition, cybersecurity training and swift regulatory responses are crucial to ensuring that institutions are better equipped to handle the evolving threat landscape.

Phishing attacks on MFIs are not unique to Kenya, India, and Nigeria. Similar incidents have been reported in other developing regions, including Latin America and Southeast Asia, where the reliance on digital platforms and limited cybersecurity infrastructure makes MFIs vulnerable (International Monetary Fund [IMF], 2023; Tech For Good Institute, 2023). For instance, a 2021 phishing attack on a microfinance institution in Brazil used WhatsApp messages containing malicious links to compromise client accounts (Shafiq et al., 2020). These cases underscore the global nature of phishing threats to MFIs, revealing common vulnerabilities such as inadequate cybersecurity training, a lack of ML-based defenses, and the absence of multi-factor authentication.

The comparison also highlights the disparity in response times and regulatory measures. In European MFIs, the adoption of machine-learning-based anomaly detection systems, such as GRUs, has proven effective in mitigating phishing attempts and reducing attack success rates (Paraschiv et al., 2021). By identifying suspicious behaviors and patterns in real-time, these systems offer an additional layer of security that MFIs in developing economies should consider

adopting. Learning from global best practices could enable MFIs in Kenya, India, and Nigeria to better protect their clients and operations from phishing threats.

3.6 Theoretical framework

This study draws upon three complementary theories—the Socio-Technical Systems Theory (STS), Institutional Theory, and the Unified Theory of Acceptance and Use of Technology (UTAUT). Together, these frameworks form a cohesive analytical foundation for exploring how technological systems, institutional structures, and human behavior collectively influence cybersecurity resilience and the adoption of artificial intelligence (AI) within digital microfinance institutions (MFIs). Although each theory provides valuable insights into specific dimensions of cybersecurity and technology adoption, none independently explains the interaction between AI-driven phishing detection, organizational governance, regulatory compliance, and user adoption within digital MFIs. Consequently, this study extended the Socio-Technical Systems Theory by explicitly incorporating governance as an integral subsystem comprising regulatory frameworks, AI governance, compliance, and institutional oversight. This extension provided a more comprehensive theoretical lens for explaining cybersecurity resilience and informed the design of the proposed AI-driven adaptive cybersecurity framework.

3.6.1 Socio-Technical Systems Theory

The socio-technical systems theory proposes that an organization's effectiveness depends on the interplay between its social and technical components (Baashar et al., 2021). In this study, socio-technical systems theory is applied to interpret phishing vulnerabilities not as isolated technological failures but as outcomes of socio-technical imbalances. The theory underscores that cybersecurity performance is shaped by the degree of harmony among human behavior, institutional culture, and the technological infrastructure that supports daily operations.

Accordingly, this research views phishing resilience as emerging from coordinated interactions among trained staff, organizational norms, and adaptive AI-based detection tools.

Recent studies extend the socio-technical systems theory perspective to cybersecurity, emphasizing the importance of balancing automated systems with meaningful human oversight in AI-driven security processes (Dwivedi et al., 2021; Lundberg et al., 2023). This insight informed the inclusion of both human-centered factors—such as user awareness, training, and trust—and technical attributes —such as model adaptability and detection accuracy —in the research design. However, traditional socio-technical systems theory approaches have often been criticized for assuming stable environments and underestimating power imbalances, particularly in resource-constrained settings (Baashar et al., 2021). To address this limitation, the current study places socio-technical interaction within the practical realities of MFIs in developing economies, where limited resources, human capacity, and technology readiness are closely intertwined.

Furthermore, this study extends the traditional Socio-Technical Systems Theory by explicitly incorporating governance as an integral subsystem comprising regulatory frameworks, AI governance, compliance mechanisms, and institutional oversight. While conventional STS applications primarily emphasize the interaction between social and technical subsystems, the findings of this study suggest that governance also plays a critical role in shaping cybersecurity resilience and the effective adoption of AI within digital MFIs. Integrating governance into the STS perspective therefore provided a more comprehensive theoretical foundation for the proposed AI-driven adaptive cybersecurity framework.

3.6.2 Institutional Theory

Institutional Theory complements the socio-technical systems framework by explaining how organizational behavior is molded by regulatory, normative, and cognitive pressures within

broader institutional environments (Mutua, 2022). This perspective offers a macro-level lens for understanding how laws, industry norms, and compliance requirements shape cybersecurity adoption among MFIs. In this study, Institutional Theory helps clarify the often-noted gap between formal compliance with cybersecurity regulations and the actual implementation of AI-based defense systems. This disconnect is frequently attributed to weak enforcement mechanisms, limited institutional capacity, and symbolic rather than substantive compliance (Mbarika et al., 2021; Papagiannidis, 2025).

The theory further explains why many MFIs in developing regions selectively adopt AI solutions, depending on available resources, institutional incentives, and perceived legitimacy. Institutional isomorphism—where organizations imitate global standards such as ISO/IEC 27001 or NIST CSF to gain credibility—provides a useful lens for understanding these behaviors (Veale & Borgesius, 2021). Nonetheless, critics have argued that the theory's deterministic tone downplays organizational agency and innovation, particularly in low-resource contexts (Mutua, 2022). To account for this, the current study integrates Institutional Theory with socio-technical systems theory, showing how institutional pressures shape socio-technical alignment, for instance, how regulatory expectations influence staff behavior, training priorities, and the configuration of AI-based cybersecurity systems in MFIs. Institutional Theory also informed the governance subsystem incorporated into the extended Socio-Technical Systems perspective by highlighting the influence of regulatory frameworks, institutional oversight, compliance requirements, and AI governance on organizational cybersecurity resilience. Consequently, governance was conceptualized not merely as an external institutional influence but as an integral component that interacts with the social and technical subsystems to support the effective implementation of AI-driven cybersecurity within digital MFIs.

3.6.3 Unified Theory of Acceptance and Use of Technology (UTAUT)

At the micro level, the Unified Theory of Acceptance and Use of Technology (UTAUT) provide valuable insights into the behavioral aspects influencing the uptake of AI-powered phishing detection tools. The theory suggests that users' intention to use technology—and their actual use—depends on factors such as performance expectancy, effort expectancy, social influence, and facilitating conditions (Venkatesh et al., 2022). Within this research, UTAUT informs the human dimension of the socio-technical framework by explaining how perceptions of usefulness, ease of use, and trust in AI outputs shape employees' willingness to engage with GRU-based phishing detection systems.

This theoretical lens was particularly relevant during the validation phase, where variables related to human–AI interaction—such as perceived interpretability, system transparency, and user confidence—were examined alongside technical performance indicators. While UTAUT effectively captures behavioral determinants of technology adoption, it has been critiqued for limited contextual flexibility in highly regulated or low-resource settings (Venkatesh et al., 2022). To mitigate this, the current study incorporates UTAUT within the broader socio-technical and institutional frameworks, acknowledging that individual technology acceptance is strongly influenced by organizational culture, management support, and institutional constraints. Accordingly, UTAUT complemented the extended Socio-Technical Systems perspective by explaining how user perceptions and behavioral intentions influenced the successful adoption of AI-driven phishing detection technologies within the governance, organizational, and technical environment of digital MFIs. Integrating UTAUT with the socio-technical and institutional perspectives therefore enabled the study to examine cybersecurity resilience from technological,

organizational, governance, and user adoption dimensions rather than relying on a single theoretical perspective.

Accordingly, UTAUT complemented the extended Socio-Technical Systems perspective by explaining how user perceptions and behavioral intentions influenced the successful adoption of AI-driven phishing detection technologies within the governance, organizational, and technical environment of digital MFIs. Integrating UTAUT with the socio-technical and institutional perspectives therefore enabled the study to examine cybersecurity resilience from technological, organizational, governance, and user adoption dimensions rather than relying on a single theoretical perspective.

3.6.4 Justification of the Selected Theories

The development of an AI-driven framework for enhancing phishing resilience in digital microfinance institutions requires an understanding of the problem from multiple perspectives. Phishing resilience is not influenced by technology alone; it is shaped by how users interact with security solutions, how institutions integrate cybersecurity practices, and how intelligent systems respond to continuously changing cyber threats. For this reason, this study adopts an integrated theoretical perspective by combining the Unified Theory of Acceptance and Use of Technology (UTAUT), Socio-Technical Systems (STS) Theory, and Adaptive Systems Theory.

The choice of these three theories was informed by the limitations of applying a single theoretical perspective to explain AI-driven cybersecurity resilience. UTAUT primarily focuses on the human dimension of technology adoption by explaining factors that influence users' acceptance and utilization of technological solutions. In the context of digital microfinance institutions, the effectiveness of an AI-based phishing detection system depends not only on its detection capability

but also on whether employees and other users are willing and able to interact with the system effectively. UTAUT therefore helps explain the behavioral factors that influence the successful adoption and use of AI-enabled cybersecurity mechanisms.

However, user acceptance alone does not guarantee cybersecurity effectiveness. AI-driven security solutions must operate within an organizational environment where policies, procedures, human expertise, and technological infrastructure interact. This study therefore incorporates Socio-Technical Systems Theory to explain the relationship between the technical components of cybersecurity solutions and the social and organizational structures that support their implementation. The theory recognizes that cybersecurity resilience emerges from the alignment between people, processes, technology, and governance mechanisms rather than from technological capability alone.

Furthermore, cybersecurity threats are constantly evolving, requiring security systems that can learn and adjust to new attack patterns. Traditional approaches that rely on fixed detection rules may struggle to address emerging phishing techniques. Adaptive Systems Theory was therefore included to explain the capability of AI-driven systems to continuously learn from new information, adjust their behavior, and improve detection performance over time. This perspective aligns with the objective of developing an adaptive neural network framework capable of responding to changing phishing threats.

The integration of the three theories provides a more comprehensive explanation of phishing resilience than any individual theory could provide. UTAUT addresses the behavioral factors influencing interaction with AI security technologies, STS Theory explains how such technologies are embedded within institutional environments, while Adaptive Systems Theory explains the ability of AI mechanisms to evolve in response to changing threats. Together, these perspectives

provide an integrated understanding of cybersecurity resilience across the individual, organizational, and technological levels.

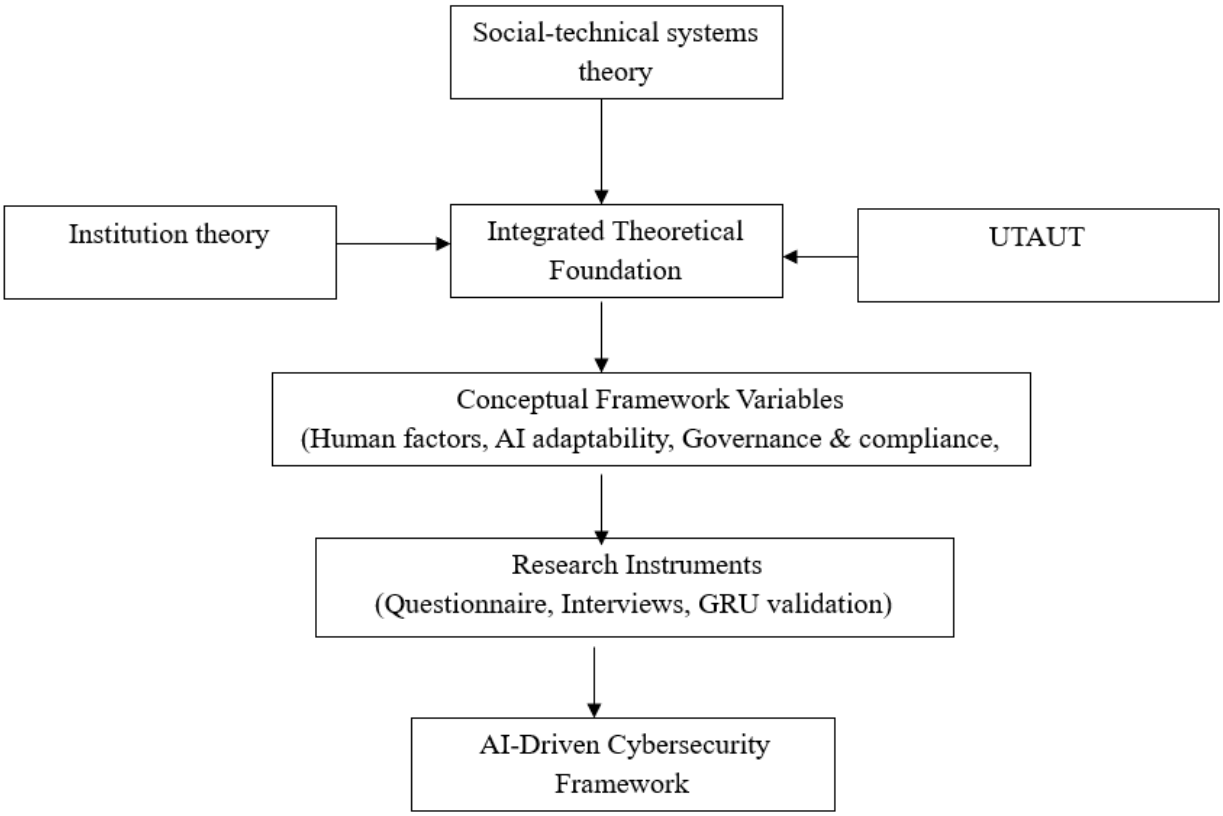
Therefore, the theoretical foundation of this study is based on the premise that effective phishing resilience in digital microfinance institutions requires more than accurate threat detection. It requires users who can effectively engage with security technologies, institutions that can successfully integrate cybersecurity practices, and intelligent systems capable of adapting to emerging threats. The integration of UTAUT, STS Theory, and Adaptive Systems Theory provides the basis for developing an AI-driven adaptive neural network framework that is technically effective, contextually appropriate, and sustainable within the microfinance environment.

3.6.5 Integration and Application to Research Design

The integration of these three theoretical perspectives offers a holistic analytical structure for the study. The Socio-Technical Systems Theory identifies how human and technical subsystems interact and where balance is required; Institutional Theory situates these dynamics within the governance and compliance landscape; and UTAUT explains how individual behaviors and perceptions influence technology adoption. Together, they informed the mixed-methods design of the research, guiding the development of questionnaires, interview protocols, and validation metrics for the GRU-based phishing detection model.

Figure 2.1 illustrates how Socio-Technical Systems Theory, Institutional Theory, and UTAUT were integrated to provide the theoretical foundation for this study. The figure shows the complementary roles of the three theories in explaining the technological, organizational, regulatory, and

behavioral dimensions of AI-driven cybersecurity within digital MFIs.



Source: Author (2025).

Figure 3.1 Integrated Theoretical Framework for the Study

Table 2.5 operationalizes this theoretical integration by mapping the core constructs of each theory to the study variables, research objectives, and data collection instruments. Rather than functioning independently, the three theories complemented one another by explaining the technological, behavioral, organizational, and regulatory dimensions of AI-driven cybersecurity in digital MFIs.

This mapping translated the theoretical concepts into measurable variables that informed the conceptual framework, questionnaire design, interview protocols, and empirical research process.

Table 3.5 Operationalization of the Integrated Theoretical Framework

Theory	**Core Constructs**	**Study Variables**	**Research Objective(s)**	**Data Collection Instrument**
Socio-Technical Systems (STS) Theory	Interaction between people, technology and organizational processes	Cybersecurity awareness, AI adaptability, communication processes, organizational support	Objectives 1, 2 and 4	Questionnaire, interviews and framework validation
Institutional Theory	Governance, regulatory compliance, organizational legitimacy and institutional readiness	Regulatory compliance, cybersecurity governance, organizational policies	Objectives 1 and 2	Questionnaire and document analysis
Unified Theory of Acceptance and Use of Technology (UTAUT)	Performance expectancy, effort expectancy, facilitating conditions and behavioral intention	AI acceptance, user adoption, behavioral intention to use AI-driven cybersecurity solutions	Objectives 2 and 4	Questionnaire and pilot validation
Integrated Theoretical Perspective	Interaction of socio-technical, institutional and behavioral factors	Cybersecurity resilience of digital MFIs	Overall research objective	Conceptual framework and integrated framework evaluation

Source: Author (2025).

As shown in Table 2.5, each theoretical perspective informed a distinct aspect of the study while collectively supporting the overall research design. Specifically, socio-technical systems theory informed the identification of core socio-technical variables such as user awareness, AI adaptability, and communication flow; Institutional Theory guided the examination of regulatory and policy frameworks influencing AI implementation; and UTAUT contributed to the development of behavioral constructs tested during pilot validation. The questionnaire items were therefore organized to capture these theoretical constructs through measurable variables relating to human behavior, organizational preparedness, governance practices, and AI adoption. Similarly, the conceptual framework integrated these constructs to explain how technological capabilities, user behavior, and institutional factors collectively influence cybersecurity resilience in digital MFIs.

This multi-theoretical integration provided methodological consistency and deepened the interpretation of both qualitative and quantitative findings.

In summary, by integrating socio-technical, institutional, and behavioral perspectives, this study establishes a comprehensive framework for understanding how AI-driven cybersecurity systems evolve in the microfinance sector. The model moves beyond a purely technological view of phishing detection by placing resilience within broader human, organizational, and regulatory contexts. The integration of the three theories therefore provided a coherent theoretical foundation that linked the conceptual framework, research instruments, and empirical analysis, ensuring that each stage of the study was grounded in complementary theoretical perspectives rather than a single explanatory lens.

Ultimately, this integrated framework advances theoretical understanding of digital security in low-resource financial ecosystems and provides a conceptual foundation for designing sustainable, transparent, and context-aware AI-based cybersecurity solutions.

3.7 Gaps Identified in the Literature

The integration of artificial intelligence-driven technologies, such as GRUs, into MFI's cybersecurity systems presents significant advantages and challenges. GRUs are an effective tool for phishing detection because they can process and analyze large volumes of data in real time, identify anomalies in phishing activity, and adapt to evolving phishing threats. However, to guarantee a successful deployment, the issues of interoperability, data accessibility, legal compliance, and budgetary limitations must be properly addressed. MFIs may strengthen their defenses against phishing attempts and enhance their overall cybersecurity posture in an increasingly digital industry by tackling these issues. However, while existing studies acknowledge these challenges, they largely fail to provide context-specific, empirically validated solutions tailored to the operational realities of MFIs, thereby revealing a critical gap in context and implementation.

The literature analysis identifies potential cybersecurity opportunities and challenges for microfinance institutions (MFIs). These include the growing threat that AI-driven phishing assaults, social engineering schemes, and advanced persistent threats (APT) pose to MFIs. MFIs are particularly susceptible to contemporary cyber threats because, compared with larger financial institutions, they lack cybersecurity expertise, use outdated technology, and have limited resources. Despite their fundamental role, traditional cybersecurity solutions are no longer sufficient to counter these intricate and dynamic attacks. Because of this, MFIs must quickly switch to proactive defenses that leverage cutting-edge technology. Nevertheless, the literature predominantly

generalizes findings from larger financial institutions, with limited focused investigation into how these threats manifest uniquely within MFIs, thereby creating a significant contextual research gap.

The ability of AI-powered technologies, particularly Gated Recurrent Units (GRUs), to improve phishing detection in MFIs is a key conclusion in the literature. GRUs are particularly good at analyzing massive amounts of sequential data in real time, spotting minute patterns and anomalies, and adapting to keep up with phishing attacks. This makes it easier for MFIs to switch from active to proactive defense techniques in cybersecurity. Regulatory compliance, budgetary constraints, data availability for AI model training, interoperability with existing systems, and other issues are impeding the deployment of GRUs and other AI-driven solutions. The successful integration of AI into the MFI cybersecurity system requires overcoming these challenges. However, there is insufficient research on the design, scalability, and cost-effectiveness of GRU-based phishing detection models specifically adapted to the resource constraints and infrastructural limitations of MFIs, highlighting a clear methodological and practical gap.

The inadequate attention paid to countering AI-driven threats and emerging social engineering techniques is another significant weakness in today's cybersecurity frameworks. Attackers are increasingly using AI to customize their phishing attempts, making it harder for traditional approaches to detect them. To identify and mitigate increasingly complex threats, artificial intelligence (AI)-based solutions, such as machine learning-powered threat detection and behavioral analytics, are crucial (Tian et al., 2021). These systems monitor for behavioral anomalies, such as unusual login times or unfamiliar devices, and send early alerts to reduce the risk of a breach. Unfortunately, due to resource constraints, many MFIs still find it difficult to deploy these cutting-edge technologies. Furthermore, existing studies provide limited empirical

evaluation of the effectiveness of AI-driven countermeasures against AI-powered phishing attacks, particularly within low-resource institutional environments, thereby indicating an evidence gap.

Collaborative cybersecurity models provide a workable and affordable solution to these drawbacks. MFIs can reduce costs and gain access to cutting-edge technology by outsourcing critical cybersecurity services, such as threat detection, incident response, and vulnerability management, to specialized providers via models like Cybersecurity-as-a-Service (CAAS) (Shafiq et al., 2020). MFIs can create shared cybersecurity frameworks that strengthen collective defenses and reduce individual financial costs by pooling their resources. To improve their overall cybersecurity posture without going over budget, a group of MFIs in East Africa, for example, might work together to improve centralized threat detection systems (Ali et al., 2023). Furthermore, governance frameworks like ISO/IEC 27001 and NIST should be modified to meet the requirements of MFIs, emphasizing fundamental security measures such as data encryption and multi-factor authentication (MFA, while permitting the gradual adoption of increasingly sophisticated technologies. However, there is a lack of longitudinal and comparative studies that assess the sustainability, scalability, and governance implications of collaborative cybersecurity models across diverse MFI contexts, revealing a strategic and policy gap.

In addition, the literature reveals a persistent human-centric gap, with limited attention to the effectiveness of continuous cybersecurity awareness training and behavioral interventions in reducing phishing susceptibility among MFI employees. Similarly, there is insufficient research on post-attack recovery mechanisms, particularly strategies for rebuilding client trust and ensuring institutional resilience following phishing incidents. Furthermore, regulatory and compliance studies inadequately address how MFIs with constrained resources can practically integrate AI-

driven cybersecurity solutions while maintaining compliance with evolving legal and ethical standards, thereby highlighting a gap in regulatory implementation.

Consistent with the systematic literature review described in Section 3.5, the research gaps presented in Table 2.6 were derived through a critical synthesis of the reviewed studies. The table brings together the key technological, human, organizational, and regulatory gaps identified in the literature, highlighting areas where existing knowledge remains limited and providing the basis for the proposed research framework. Table 2.6 summarizes the research gaps identified across the reviewed literature and illustrates the technological, human, organizational, and regulatory issues that informed the development of the proposed research framework.

Table 3.6 Summary of gaps in the literature on phishing attacks on digital MFIs

Area of literature	**Key Findings**	**Gaps Identified**
Trends of phishing attacks in MFIs	Compared to larger financial institutions, MFIs have been overwhelmed by evolving phishing attacks due to their weaker defense systems and overreliance on traditional cybersecurity measures	Lack of in-depth research on vulnerabilities unique to MFIs that differentiate them from larger financial institutions
phishing attacks techniques	Spear phishing and whaling attacks are continuously targeting MFI executives	Lack of cost-effective, scalable cybersecurity solutions uniquely tailored to the technical, financial, and

		operational constraints of MFIs
Social Engineering and sophisticated AI-powered phishing	Sophisticated AI and advanced social engineering are being used to customize manipulative phishing attacks, tricking employees into revealing sensitive information	Poor adoption of AI-driven solutions to counter evolving AI-driven phishing attacks
Vulnerability of employees to phishing attacks	Phishing attacks exploit human weaknesses, and this overrides the technical safeguards, thereby exposing the resource-constrained MFIs	Lack of sufficient research on the impact of cybersecurity training on employees in identifying and mitigating phishing attacks in MFIs
Financial and reputational impact of phishing on MFIs	Phishing attacks have led to huge financial losses, data breaches, eroding client trust, resulting in operational delays and expensive compliance issues	Insufficient research on workable long-term strategies and solutions for restoring client trust after phishing attacks in MFIs
Compliance and regulatory challenges	MFIs are struggling to meet strict compliance	Limited research on how MFIs can achieve regulatory

	requirements and are being exposed to phishing attacks	compliance with limited resources, especially AI-driven solutions
Emerging trends in phishing attacks (deepfake and AI-powered phishing attacks)	In the ever-evolving landscape of phishing attacks, MFIs are now targeted by deep fakes and AI-powered attacks. This has proved difficult for such institutions to safeguard themselves against data breaches, thereby compromising the integrity	There is a lack of sufficient research on how resource-constrained MFI can effectively implement AI-driven solutions to combat the ever-evolving AI-powered threats

Source: Compiled by the researcher from the systematic literature review undertaken for this study

Overall, the literature review shows that although there has been substantial progress in understanding phishing threats and AI-based detection methods, a key research gap remains: the development of an integrated, cost-effective, and context-specific framework for AI-based phishing detection in MFIs. This framework should consider technological, human, and regulatory factors. This study aims to fill that gap by focusing on GRU-based methods within the specific operational environment of MFIs.

3.8 Conceptual Framework

The study's conceptual framework demonstrates how AI-enabled cybersecurity systems and phishing detection in digital microfinance institutions (MFIs) are interrelated. The framework is based on the study objectives and the literature review and is grounded in existing theoretical perspectives, such as the Socio-Technical Systems Theory, the Institutional Theory, and the Unified Theory of Acceptance and Use of Technology (UTAUT). The combination of these theories provides a way to explain the relationships among technological systems, the organizational environment, and user behaviors in determining system performance and adoption.

The framework hypothesizes that GRU model design, dataset quality, and AI-based cybersecurity integration strategy influence phishing detection effectiveness both directly and indirectly through feature representation quality and model learning capability. These relationships are further moderated by organizational, regulatory, and human factors that characterize the operating environment of digital MFIs.

The framework theorizes the implementation of the AI-driven GRU models, the quality of the data set, and the strategy to integrate cybersecurity as the independent variables affecting phishing detection effectiveness. GRU model design is operationalized through architectural characteristics such as hidden units, learning parameters, computational efficiency, and classification performance. Dataset quality is assessed using dataset size, diversity, multilingual coverage, class balance, and relevance to phishing detection. AI-based cybersecurity integration strategy refers to the extent to which AI-driven phishing detection is embedded within existing cybersecurity infrastructure through real-time threat detection, interoperability with existing security systems, automated threat response, and operational deployment within digital MFIs. GRU model architecture is modeled in terms of model accuracy, recall, precision as well as computational

efficiency. The quality of datasets is assessed by dataset size, diversity, multilingual coverage, and phishing relevance, whereas cybersecurity integration strategy can be considered the degree of AI embedding into the current detection systems, including real-time processing facilities and the ability to be integrated with the existing systems.

Phishing detection effectiveness is the dependent variable and operationalized by measuring the accuracy of detection, false positive rate, F1-score, and the ability of the system to handle changing phishing threats in real time.

The framework adds model learning capability and the quality of feature extraction as mediating variables to provide more explanatory power. These variables determine the process by which the independent variables affect the results of detection. In particular, high-quality datasets and GRU architectures improve feature extraction and learning, thereby improving detection performance.

Moderating variables also affect the relationship in the framework, such as resource constraints in MFIs (e.g., computing capacity and technical skills), regulatory and governance structures (e.g., compliance needs and AI governance systems), and user factors (e.g., staff awareness and vulnerability to social engineering attacks). These moderating variables affect the strength and direction of the relationships between the independent variables, mediating processes, and the dependent outcome.

As illustrated in Figure 2.2, the conceptual framework presents a causal pathway linking the independent, mediating, moderating, and dependent variables. The solid arrows represent the direct influence of the independent variables on the mediating variables and, subsequently, on phishing detection effectiveness. The dashed arrows indicate the moderating influence of organizational, regulatory, and human factors on these relationships. This arrangement illustrates

that the effectiveness of AI-driven phishing detection depends not only on the technical performance of the GRU model but also on the institutional environment in which it is deployed.

It is an interactive and interdependent system whereby the performance of technologies is influenced not just by model design and data quality but by organizational and contextual influences. The framework thus takes a socio-technical approach whereby technological efficiency is not enough without the inclusion of institutional and human factors. It is also consistent with the Institutional Theory because it emphasizes the contribution of regulatory and organizational forces to the shaping of AI acceptance, and with the UTAUT because it integrates factors that are related to users and can affect the usage of systems and their effectiveness.

The conceptual framework therefore integrates the three underpinning theories into a single analytical model that guided the selection of study variables, informed the design of the research instruments, and provided the basis for examining the technological, organizational, behavioral, and regulatory factors influencing phishing detection in digital MFIs. Consequently, the framework served as the foundation for the empirical investigation and the development of the proposed AI-driven cybersecurity framework.

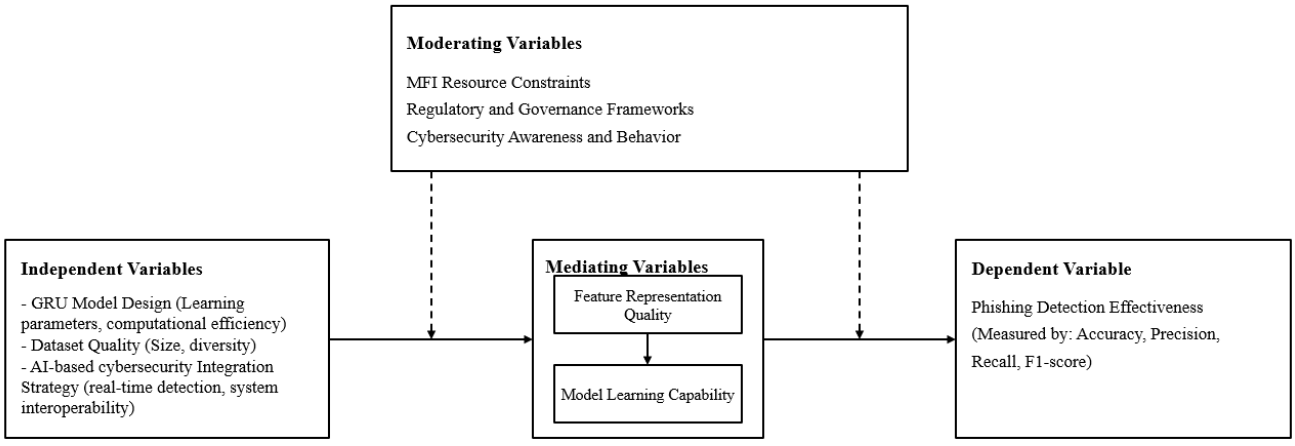


Figure 3.2 Conceptual Framework for AI-Driven Phishing Detection

3.9 Summary

This chapter reviewed the existing literature on phishing attacks, cybersecurity vulnerabilities, artificial intelligence, and governance frameworks relevant to digital microfinance institutions (MFIs). The review showed that phishing attacks had become increasingly sophisticated through the use of social engineering and AI-enabled techniques, while many MFIs continued to face heightened cybersecurity risks because of limited resources, outdated technological infrastructure, and insufficient institutional capacity. As a result, conventional cybersecurity measures were found to be inadequate for addressing the rapidly evolving threat landscape.

The review also showed that, although artificial intelligence offered considerable potential for strengthening phishing detection and cybersecurity, existing studies had paid limited attention to developing integrated and context-specific solutions that reflected the unique operating environment of MFIs. In addition, important technological, human, organizational, and regulatory gaps remained, limiting the effective adoption of AI-driven cybersecurity approaches within resource-constrained institutions.

Overall, the literature provided a clear understanding of the current state of knowledge while highlighting the areas that required further investigation. The identified gaps justified the need for an integrated and context-sensitive cybersecurity framework for digital MFIs and informed the development of the conceptual framework presented in the next section.

CHAPTER 4

RESEARCH METHODOLOGY

4.1 Introduction

This chapter describes the research approach applied to explore the phishing vulnerabilities in Nairobi, Kenya's microfinance institutions (MFIs). The study investigated the contextual cybersecurity procedures within these MFIs and the technical effectiveness of an AI-driven Gated Recurrent Unit (GRU) neural network in identifying phishing attempts. The methodology employed a mixed-methods approach to thoroughly explain phishing vulnerabilities and institutional responses, combining quantitative and qualitative analyses. The chapter covers research design, population and sampling, data collection and analysis methods, ethical concerns, validity and reliability, and the study's limitations and delimitations.

4.2 Research philosophy

To address the socio-technical aspects of phishing vulnerabilities in microfinance institutions (MFIs) and to develop an AI-driven solution, this study is firmly rooted in the research philosophy of critical realism and its epistemological paradigm. This is because it acknowledges that phishing threats and AI model performance exist as objective phenomena while recognizing that their implementation and effectiveness are influenced by organizational, social, and technological contexts. While acknowledging objective truths, such as the prevalence of phishing vulnerabilities and the technical effectiveness of AI frameworks, critical realism also recognizes that institutional, social, and technological constructs shape how humans perceive this reality (Elder-Vass, 2022). This dual viewpoint is ideal for the mixed-methods technique used in this study, which combines quantitative evaluation of the GRU neural network model with qualitative investigation of

institutional behaviors. The philosophy was therefore appropriate because the study sought not only to measure the predictive performance of the proposed GRU model but also to explain how organizational governance, institutional practices, and user behavior influenced the adoption and effectiveness of AI-driven cybersecurity within digital MFIs.

The choice of epistemology philosophy was especially justified by the study's need to reconcile empirical evidence and interpretive knowledge. As a branch of knowledge theory, epistemology provides the basis for understanding how legitimate knowledge about phishing threats and AI-related responses may be created. The study adopted a critical realist epistemological approach and recognizes that phishing and AI model outputs can be observed, but their meaning is mediated by human interpretation, institutional policies, and socio-technical contexts. It is within this paradigm that the research can justify AI-based detection on the basis of measurable consequences (positivist orientation) and, at the same time, challenge institutional governance practices by exploring them through interpretive inquiry. This integration aligns with current information systems research, which emphasizes the integration of positivist and interpretivist epistemologies to address complex socio-technical problems (Mingers et al., 2020; Zachariadis et al., 2022). Additionally, this methodological coherence allows the quantitative accuracy of the GRU neural network to be complemented by qualitative depth, a key requirement of cybersecurity studies in digital microfinance (Myers & Klein, 2021; Rieder et al., 2025). Accordingly, critical realism provided an appropriate philosophical foundation for integrating quantitative model evaluation with qualitative inquiry, enabling the study to explain not only whether the proposed framework was effective but also why its implementation and adoption were shaped by the socio-technical and institutional realities of digital MFIs. By so doing, the epistemological paradigm offers the

philosophical rationale and methodological framework needed to question the multidimensional realities of phishing in MFIs.

4.3 Research Design

To thoroughly examine the two aspects of technological efficacy and contextual cybersecurity within MFIs, the study employed a mixed-methods approach combining quantitative and qualitative methods (Johnson et al., 2024). The quantitative component mainly focused on providing empirical data on the GRU model's accuracy, precision, and adaptability in phishing detection. The qualitative component investigated how operational, human, and institutional elements influence cybersecurity vulnerabilities and behaviors. A mixed-methods approach was selected because neither quantitative nor qualitative methods alone could comprehensively address the study objectives. While the quantitative approach measured the technical performance of the GRU model, the qualitative approach provided contextual insights into organizational practices, user behavior, and governance factors influencing cybersecurity resilience. Integrating both approaches enabled a more comprehensive understanding of phishing threats and supported the development and validation of an adaptive AI-driven framework suited to digital MFIs. Using these strategies, the study sought to comprehensively understand the challenges and possible solutions in resource-constrained cybersecurity environments for MFIs. The justification and research design for each research objective are listed below

Specific Objective (i): To map cybersecurity vulnerabilities facing digital microfinance Institutions that can be addressed through AI-driven solutions.

This objective was addressed using a qualitative research design. Detailed information on the elements causing cybersecurity risks was obtained through semi-structured interviews with MFI managers and IT staff, as well as a thorough review of operational and policy documentation. Because it enabled a deeper understanding of intricate, context-specific concerns that cannot be adequately captured by quantitative data alone, this technique is ideal. The results informed the process of identifying vulnerabilities and the possibility of AI-driven solutions.

Specific Objective (ii): To assess the effectiveness of existing cybersecurity frameworks in mitigating phishing threats within digital microfinance institutions.

To achieve this objective, a quantitative evaluative design was employed. The performance of the current cybersecurity frameworks was assessed using data gathered from structured questionnaires and document analysis. While inferential analysis, such as regression modeling, evaluated the frameworks' efficacy in reducing phishing incidents, descriptive statistics provided a general picture of adoption rates. This approach guaranteed an unbiased, fact-based evaluation of current frameworks and their limitations.

Specific Objective (iii): To develop a neural network AI-driven adaptive framework, using GRU neural network models, for improving phishing threat detection in digital microfinance Institutions.

To accomplish this specific objective, a design-and-development research methodology was used. Labeled phishing datasets were used to develop and train the GRU-based adaptive framework, which included iterative testing and optimization. This concept was well-suited to developing a strong, flexible framework to address the challenges posed by phishing attacks in MFIs. Ensuring

model correctness, precision, and responsiveness to changing threat patterns was the main goal of the development process.

Specific Objective (iv): To validate the effectiveness of the developed adaptive AI-driven model in mitigating phishing threats within digital microfinance Institutions.

This objective validated the created model using a quasi-experimental design. Real-world scenarios and benchmark phishing datasets supplied by partner institutions were used to evaluate the GRU model's performance. To assess the model's effectiveness, key evaluation metrics, including accuracy, recall, precision, and F1 score, were evaluated. This method guarantees thorough validation while preserving adaptability when integrating the framework into practical uses.

Table 3.1 provides a clear overview of the research designs used throughout the thesis chapters, showing how each aligns with particular research questions. It reflects a thoughtful and systematic approach, ensuring that the chosen methods are well-suited to the specific aims and nature of each question.

Table 4.1: Research designs and research questions addressed in the different chapters of the thesis

Objective	**Research question**	**Research approach used**	**Data**	**Analytical tool**	**Implementation Chapter (s)**
To map cybersecurity vulnerabilities	How can the key cybersecurity	Systematic literature review	Industry reports, case studies, peer-	Systematic literature review,	Chapter two (Literature Review),

facing digital microfinance Institutions that can be addressed through AI-driven solutions.	vulnerabilities in digital microfinance Institutions be identified and mitigated using AI-driven solutions?		reviewed journal articles	Thematic analysis, Descriptive statistics	Chapter three (Methodology)
To assess the effectiveness of existing cybersecurity frameworks in mitigating phishing threats within digital microfinance institutions.	How effective are the existing cybersecurity frameworks and policies in mitigating evolving phishing threats facing microfinance Institutions?	Mixed-methods approach	Regulatory cybersecurity frameworks, case studies, policy documents, and primary data collected through structured questionnaires	Content analysis, Comparative evaluation, descriptive statistics	Chapter two (framework review), Chapter three (data collection), chapter four (empirical findings)

To develop an AI-driven, adaptive framework using GRU neural network models to improve phishing threat mitigation in digital microfinance Institutions.	How can GRU neural network models be optimized to accurately detect, prevent, and predict phishing threats within digital microfinance Institutions?	Model training and development	Phishing datasets – Kaggle, institutional logs	Python Libraries – pandas, TensorFlow, Keras. Performance metrics (accuracy, precision, recall, F1 score	Chapter three (model development methodology), chapter 4(framework development)
To validate the effectiveness of the developed AI-driven adaptive framework in	How effective is the newly developed AI-driven framework in improving the detection and mitigation of phishing threats	Model testing and evaluation	Phishing datasets – test and validation data	ROC curves, confusion matrices, and statistical comparison	Chapter three (validation methodology, Chapter four (framework validation)

mitigating phishing threats within digital microfinance Institutions.	compared to existing methods within digital microfinance Institutions?				

Source: Author (2025).

4.3.1 Research Strategy

The study combined experimental testing with a case study approach. The case study was a suitable approach for a thorough examination of cybersecurity vulnerabilities within MFIs (Schoch, 2020). The case study enabled a thorough analysis of institutional procedures and practices, as well as the specific challenges faced by MFIs in Nairobi. An array of divergent viewpoints and perspectives from frontline staff to IT personnel were captured through surveys and interviews conducted at several institutions for the case study.

Experimental testing, on the other hand, is the systematic evaluation of a framework or model. This is achieved in a controlled environment to help determine the performance, adaptability, and reliability of the given framework or model (Kamiri et al., 2021). The objective of the experimental component was to evaluate the GRU model's performance in detecting phishing attempts. The testing was carried out in a controlled environment using simulated phishing data. The model's technical efficacy in the MFI context was empirically demonstrated through experiments measuring the model’s detection accuracy, false-positive rates, and responsiveness to new phishing techniques.

4.3.2 Justification for Chosen Design

The experimental design and the mixed-methods approach are an ideal choice for this study for several reasons:

The study addresses both the technical and contextual aspects of the problem by combining experimental testing with a case study, enabling a comprehensive examination of phishing detection in MFIs (Akerblad et al., 2021). Incorporating the case study component provided an in-depth understanding of the unique cybersecurity issues and organizational dynamics within MFIs, including capacity limitations and employees' behavior in response to phishing attempts.

The experimental component thoroughly evaluates the GRU model's performance and provides measurable insights about its applicability and adaptability in the ever-changing threat landscape of digital MFIs (Leavy, 2022). This method not only evaluated the efficacy of the GRU model but also produced insights that may help other MFIs with comparable limitations adapt to AI-based security frameworks, thereby enhancing the study's overall applicability and relevance.

4.4 Research Process

This section provides a thorough, organized summary of the research process used to develop the GRU-based framework for phishing mitigation. The process is structured into six consecutive steps, starting with problem definition and ending with model deployment, as shown in Figure 3.1.

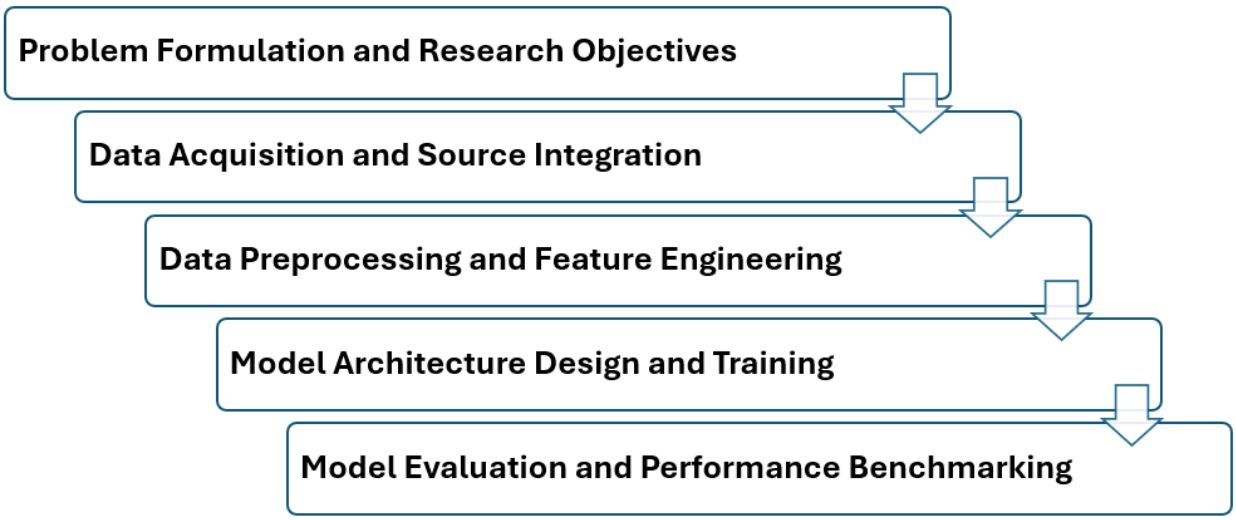


Figure 4.1: Step-by-step Research Process

4.4.1 Step 1 - Problem formulation and Research objectives

The study started with a thorough analysis of Kenya's microfinance institutions' (MFIs') vulnerabilities connected to phishing. The review emphasized the serious shortcomings of conventional detection technologies, particularly their inability to recognize evolving social engineering attempts. An adaptive AI-based detection system suited to the operational reality of MFIs was developed as a result of these findings, which also influenced the formulation of the study topic and the setting of objectives.

4.4.2 Step 2 - Data acquisition

The study used both primary and secondary data sources, as indicated in Step 2 of Figure 3.1. Interviews and surveys with IT officers and other IT frontline staff from MFIs in Nairobi were used to gather primary data. Among the secondary datasets are: Mendeley's Phishing Website Dataset (URL-based attacks), the CMU Enron E-mail Corpus (spam and phishing emails), and network traffic patterns from the Canadian Institute for Cybersecurity's CICIDS Dataset.

To train a reliable model capable of recognizing multi-vector phishing patterns, these datasets provided a variety of input modalities, including textual, structural, and behavioral characteristics.

4.4.3 Step 3 - Data Processing

Every dataset was preprocessed for GRU compatibility, including sequence padding, normalization, one-hot encoding of categorical features, and removal of null values. Techniques from natural language processing were used to tokenize email addresses and URLs. Data from flow-based networks were sessionized to identify patterns across time. Through feature engineering, behavioral indicators such as packet frequency, domain entropy, and redirect depth were extracted. As shown in Step 3 of Figure 3.1, this step illustrates how these activities ensure consistent input representation for the model.

4.4.4 Step 4 - Model Architecture Design and Training

Python was used with the TensorFlow, Scikit-learn, and Keras libraries to create the GRU model. The architecture included an input layer, a dense layer, a GRU block, a sigmoid output layer, and dropout layers for regularization. As shown in Step 4 of Figure 3.1, 70% of the data was used for model training, and a grid search was employed to optimize performance via hyperparameter tuning. During classification, the binary cross-entropy loss function and the Adam optimizer were used to ensure stable convergence.

4.4.5 Step 5 - Model evaluation and validation

Cross-validation techniques were used to evaluate the GRU model's validity, ensuring it generalizes across diverse phishing datasets and adapts to emerging attack vectors. The dataset was partitioned into training, validation, and testing subsets using a stratified 70:20:10 ratio. The validation subset supported model tuning, while the test subset was reserved for final performance evaluation. Several metrics were used to evaluate the model, including ROC-AUC, F1 Score,

recall, accuracy, and precision. These were chosen because they have demonstrated efficacy in phishing detection tasks using unbalanced data (Almiani et al., 2023; Islam et al., 2022). By using simulated phishing scenarios adapted from actual MFI attack narratives, the evaluation was enhanced through contextual validation, ensuring applicability to operational situations.

4.5 Population of the Study

The study population was defined for each research objective. The term "population" is used in a broader sense to refer to the different units of analysis considered at each stage of the research, including literature sources, institutional participants, and datasets. This broader definition reflects the mixed-methods design adopted in the study; whereby different units of analysis were considered for different research objectives.

For Objective 1, which relied on a systematic literature review, the study considered a body of peer-reviewed journal articles, industry reports, and regulatory cybersecurity frameworks relevant to phishing and microfinance institutions. The systematic literature review was conducted using a predefined search and study-selection procedure, and its reporting was guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). PRISMA 2020 provides a structured framework for transparently reporting the identification, screening, eligibility assessment, and inclusion of studies in systematic reviews. In this study, the PRISMA 2020 process was applied by documenting the records identified through the predefined search strategy, removal of duplicate records, screening of titles and abstracts, full-text assessment against the eligibility criteria, and final inclusion of studies in the review. A total of 1,224 studies were initially identified from a combination of major academic databases, including Web of Science and Scopus, and other relevant sources such as ScienceDirect, IEEE Xplore, BASE, CORE, and Google Scholar, to ensure broad and balanced coverage of both

technical and applied research. The figure of 1,224 represented the records returned by the predefined search strategy and was therefore not a statistically calculated sample size. A mathematical sampling formula was not applied because the purpose of the systematic literature search was to identify potentially relevant studies through predefined search terms and eligibility criteria rather than to draw a predetermined statistical sample from a known population. At the identification stage, 1,224 records were retrieved and 497 duplicate records were removed, leaving 727 unique records for screening. After removing 497 duplicate records, 727 unique studies remained and were screened based on their titles and abstracts. At the screening stage, the 727 records were assessed based on their titles and abstracts, after which 157 reports were retained for full-text assessment. From this pool, 157 studies were selected for full-text review, and ultimately 32 met the final inclusion criteria based on relevance, recency, and methodological quality. Thus, the 1,224 records represented the initial retrieval pool, while the 32 studies constituted the final body of literature included in the review. The study-selection process therefore corresponded to the PRISMA 2020 identification, screening, eligibility, and inclusion stages (Page et al., 2021). Table 3.2 summarizes the distribution of studies across the main sources at the identification stage.

Table 4.2 Distribution of Retrieved Studies by Source

Database/Source	**Number of Studies Retrieved**
Web of Science	88
Scopus	401
BASE	434
CORE	85
Google Scholar & Other Sources	216
Total	**1,224**

Although ScienceDirect and IEEE Xplore were part of the search process, their results are largely captured within broader indexing platforms such as Scopus and Web of Science, or appear within the "Google Scholar and other sources" category. For this reason, they are not listed separately in the table, as doing so would risk double-counting the same studies.

These databases were selected to ensure comprehensive coverage of high-quality literature relevant to artificial intelligence, cybersecurity, and digital financial systems. Scopus and Web of Science were used because they provide reliable, well-indexed, peer-reviewed literature. BASE and CORE helped capture open-access studies, which are especially important for emerging and context-specific research. Google Scholar was included to broaden the search and capture relevant reports, policy documents, and recent studies that may not yet be indexed elsewhere. Together, these sources ensured that the review covered both the technical aspects of AI and cybersecurity and the practical realities of digital financial systems (Dwivedi et al., 2021). Studies were included based on their relevance to the research objectives, publication quality, methodological rigor, and recency, while duplicate records and studies that did not meet the eligibility criteria were excluded.

For Objective 2, the institutional population comprised licensed microfinance institutions (SACCOs) operating in Nairobi. Within each sampled institution, one eligible respondent drawn from the Information Technology, Finance, Customer Service, Risk Management, or Digital Financial Services function participated in the study because these personnel were directly involved in cybersecurity operations. The sample size is explicitly discussed in detail in section 3.5.2.

This group was selected because they work directly with digital financial systems and are therefore most exposed to cybersecurity risks, especially phishing attacks. Their day-to-day responsibilities position them well to provide practical insights into how institutions manage cybersecurity, adopt AI tools, and respond to emerging threats. Nairobi was chosen because it is Kenya's main financial and technological hub, with a high concentration of MFIs and fintech services, making it an ideal setting for this study (Mwangi & Otieno, 2021).

For Objectives 3 and 4, the study used phishing datasets from sources such as Kaggle, comprising over 50,000 labeled instances of phishing and legitimate data to train and validate the GRU-based model. Using a dataset of this size was important for building a reliable and effective model. Larger datasets allow deep learning models such as GRU to learn patterns more accurately and reduce the risk of overfitting. In addition, combining publicly available data with institutional data helped ensure that the model reflects real-world phishing scenarios faced by microfinance institutions.

Licensed Savings and Credit Cooperative Society (SACCOs) and mobile money services like M-Pesa are among the digital microfinance institutions (MFIs) operating in Nairobi and constitute the target population. These MFIs have constrained cybersecurity resources and limited cybersecurity expertise. They handle high volumes of digital transactions, thereby becoming increasingly susceptible to phishing attacks. Nairobi was selected for the study because it is Kenya's financial capital for digital services and is home to many MFIs that serve a wide range of clients and are particularly vulnerable to cyber threats (Mwangi & Otieno, 2021).

4.5.1 Sampling Technique

This study employed purposive sampling, in which units or participants are selected based on predefined characteristics or criteria relevant to the study's objectives. This approach ensured that

the sample was particularly well-suited to address the research questions by focusing on participants with specialized knowledge, experience, or exposure to the topic under investigation (Bell & Bryman, 2023). In this study, purposive sampling was used to select a sample of MFIs in Nairobi that are recognized for their reliance on digital services and demonstrated vulnerability to phishing attacks. This ensured that the sample was highly relevant, enabling a comprehensive analysis of phishing threats and the effectiveness of the recommended AI-powered solutions. This sampling strategy provided a representative sample of Nairobi's digital MFI ecosystem, ensuring the inclusion of institutions with significant exposure to cyber-attacks. The unit of analysis for this study was participants from a variety of positions within each institution, including frontline staff, management, and IT security personnel, who were selected through stratified sampling. The study's ability to capture a wide range of viewpoints and experiences regarding phishing risks was enhanced by stratified sampling, which provided a comprehensive picture of cybersecurity practices and attitudes across institutional hierarchies.

This sampling approach was applied specifically to the study's primary data collection component (structured questionnaires), in which participants were selected based on their roles and direct involvement in cybersecurity and digital financial operations.

4.5.2 Sample Size

As shown in Equation 3.1 below, the Yamane (1967) formula was used to calculate the proper sample size for this research. With the appropriate degree of precision taken into consideration, this approach guaranteed that the chosen sample size was statistically representative of the target population (Chaokromthong, 2021)

$$n = \frac{N}{1 + N(e)^2}$$ *Equation 4.1*

Where:

n = sample size,

N= is the population size taken from the list of licensed Saccos 2024, and

e= is the level of precision.

The study, therefore, assumed a 95% confidence level and 5% precision. By applying Yamane's formula and substituting the figures, the resulting value was 122 (Equation 3.2)

$$n = \frac{N}{1 + N(e)^2} = \frac{176}{1+(176((0.05)^2))} = 122$$ *Equation 4.2*

The 176 licensed SACCOs constituted the institutional sampling frame for this study. Applying Yamane's (1967) formula yielded a sample of 122 institutions. From each selected institution, one respondent who was directly involved in cybersecurity, information technology, digital financial services, finance, customer service, or risk management was purposively selected to participate in the study. Consequently, the final sample comprised 122 respondents representing 122 licensed SACCOs. This approach ensured that each response reflected institutional cybersecurity practices while maintaining the statistical representativeness of the sampled institutions. Although the institutional sampling frame consisted of licensed SACCOs, the unit of observation was the individual employee because the study sought expert opinions from personnel directly involved in cybersecurity and digital financial service operations within each institution.

4.5.2.1 Justification for Use of Yamane's formula

For the purposes of this research, Yamane's sampling technique was chosen because it is well-suited to calculating statistically representative sample sizes for finite populations. Given the specified number of Sacco societies licensed in 2024, the method ensured that the selected sample size was adequate and effective, striking a balance between the necessary precision and available resources (Chaokromthong, 2021). Its significance was further supported by the fact that the 95% confidence level and 5% margin-of-error assumptions are consistent with accepted research procedures.

Yamane's formula's ease and simplicity of use are among its primary advantages. In contrast to other techniques, such as Cochran's formula, which require further adjustments for finite populations, Yamane's method directly incorporates the population size (N) into the computation. This feature optimizes the sampling procedure without affecting accuracy, as it eliminates the need for additional corrections.

The formula is especially useful for controlling the acceptable margin of error, as it explicitly incorporates precision (e). This reduces the risk of over- or under-sampling and ensures the sample size is adequate to accurately reflect the population's characteristics. When both thematic analysis and statistical testing are key objectives of a study, this emphasis on accuracy is especially essential.

4.5.3 Inclusion and Exclusion Criteria for Study Participants

For the first objective, studies were included if they focused on cybersecurity, phishing attacks, or AI applications in financial systems and were published in credible academic or industry sources.

Studies that were not directly relevant to microfinance institutions or lacked sufficient methodological rigor were excluded.

For the second objective, MFIs were considered for inclusion if they met the criteria of having a personnel structure that incorporated cybersecurity and operations functions, being digitally equipped, and actively participating in online transactions. Participants were selected based on their firsthand experience-with cybersecurity or digital transactions. A pre-screening questionnaire evaluating participants' roles, responsibilities, and direct interaction with cybersecurity systems or digital infrastructure was used to identify those with firsthand experience. Self-reported instances of relevant duties, time spent in such roles, and, when necessary, supporting documentation, such as job descriptions or certifications, were considered for verification.

Organizations that outsourced their cybersecurity services or lacked significant digital infrastructure were excluded from the study, as these characteristics may introduce confounding variables that distort the outcome.

4.5.4 Justification of Study Area

Nairobi is the capital and largest city of the Republic of Kenya, with an administrative area of approximately 700 km² (Nairobi City County Government, 2023). It was selected as the study area because it is Kenya’s principal financial and digital services hub, hosting a high concentration of licensed microfinance institutions (MFIs), financial technology providers, and digital payment platforms. The city also serves as the country's center for administration, commerce, and innovation, making it an appropriate setting for investigating cybersecurity challenges affecting digital financial services.

The 2019 Kenya Population and Housing Census reported a population of 4,397,073, with projections indicating continued growth (Kenya National Bureau of Statistics [KNBS], 2019a, 2019c). However, the selection of Nairobi was not based on population size or density alone. Rather, it was informed by the city's high level of digital financial activity, widespread adoption of mobile and online financial services, and concentration of institutions that increasingly rely on digital platforms for service delivery. These characteristics increase the volume of digital interactions and expand the potential attack surface for phishing and related cyber threats.

As Kenya's leading financial center, Nairobi accounts for a substantial share of formal economic activity and hosts the headquarters and operational centers of many financial institutions, including MFIs (Nairobi City County Government, 2023). The concentration of digital financial transactions and technology-enabled services provides a suitable environment for examining how phishing attacks affect resource-constrained financial institutions and how AI-driven cybersecurity solutions can be deployed to mitigate these threats.

Previous studies have similarly identified Nairobi as a major hotspot for cyber threats because of its extensive digital financial ecosystem and high reliance on internet-enabled financial services (Mwangi & Otieno, 2021). Consequently, the city provided an appropriate context for investigating phishing vulnerabilities, evaluating existing cybersecurity practices, and developing an AI-driven adaptive framework tailored to the operational realities of digital MFIs.

4.6 Data Collection Methods

The study included both primary and secondary data collection methods. Using the primary methods allows for the collection of original data to fill identified gaps and support the study comprehensively. Using secondary data allows for leveraging existing literature and knowledge.

4.6.1 Primary Data Collection Methods

Key personnel, such as frontline staff, cybersecurity officers, and IT managers, participated in in-depth, semi-structured interviews. Institutional views on phishing risks, the perceived efficacy of current security measures, and any organizational obstacles to putting advanced security frameworks into practice was covered in interview questions.

To gather information on cybersecurity training, response practices, and phishing awareness, structured questionnaires were administered to a larger sample of staff members within each organization. The purpose of surveys was to provide a reliable quantitative measure of staff attitudes, knowledge, and awareness in the cybersecurity sector.

4.6.2 Secondary Data Collection Methods

To place the primary findings in the perspective of larger industry trends, secondary data sources included industry reports, phishing incident reports, regulatory data, and case studies. Data from related research on phishing in online financial institutions was used to improve the study's theoretical foundation and strengthen its conclusions. The phishing websites, email phishing, and network traffic anomaly datasets from Kaggle were used in this study. These datasets were selected due to their various features and high-quality labeling, which are essential for training and

evaluating GRU models in phishing detection. The study's theoretical framework and results was reinforced by data from similar studies on phishing in online financial institutions

4.6.3 Development of Data Collection Tools

This section explains how the data collection instruments, specifically the structured questionnaire and the experimental datasets, were developed and aligned with the study objectives.

The structured questionnaire was designed to gather information on cybersecurity practices, staff awareness of phishing threats, and digital microfinance institutions' responses to such risks. The questionnaire was organized into sections covering demographic details, cybersecurity awareness, exposure to phishing attacks, and institutional mitigation strategies. Most of the questions were framed using Likert scales to capture respondents' perceptions and levels of awareness, while a few items collected categorical information such as role and experience.

The review included case studies, industry publications, regulatory documents, and journal papers that focus on cybersecurity and phishing vulnerabilities in digital financial institutions. These sources helped shape the questionnaire by ensuring that the questions reflected both the technical aspects of cybersecurity and the human factors that influenced phishing risk in MFIs.

The experimental component of the study focused on training and testing the GRU-based model using phishing datasets. This approach was guided by existing studies on AI-driven cybersecurity and phishing detection, which highlighted the importance of using well-labeled, diverse datasets to improve detection accuracy (Kar et al., 2024). At the same time, the questionnaire design was informed by prior research on user behavior and phishing susceptibility, which emphasized the role of human factors in cybersecurity (Oner et al., 2025).

Datasets from reputable platforms like Kaggle were analyzed to enhance model development, ensuring trustworthy data for neural network training and testing. The dataset included network traffic anomalies, phishing emails, and phishing websites because they met the GRU model's requirements for sequential data. These datasets were chosen for their applicability, variety of attributes, and reliable labeling, guaranteeing thorough training and testing for phishing detection in digital microfinance institutions. In selecting these datasets, particular attention was given to their relevance to phishing detection, the diversity of features they contained, and the quality of their labeling, as these factors were essential for effective training and evaluation of the GRU-based model within the proposed framework.

Overall, the combination of questionnaire data and experimental datasets enabled an examination of phishing risks from both human and technical perspectives, providing a more complete understanding of vulnerabilities and potential mitigation strategies in digital microfinance institutions.

4.6.4 Questionnaire and Interview Guide Instrument Validity

Pilot testing of the survey tools was conducted with a sample of 10 respondents, randomly selected to represent key positions within the study's target group. They included key operations staff, cybersecurity analysts, and IT managers. The pilot sample was considered adequate for identifying ambiguities, testing clarity, and assessing the instrument's overall usability at an early stage. The pilot sample represented approximately 8% of the calculated sample size, which is consistent with recommendations for pilot studies and was considered adequate for refining the research instruments before the main study. The validity and reliability of the survey was ascertained by this pilot. Internal consistency metrics such as Cronbach's alpha were used to assess reliability; a

threshold of at least 0.7 was used to confirm acceptable reliability, which is consistent with recent methodological literature. Hair et al. (2021) and Dorsa (2026) indicate that a Cronbach's alpha value of 0.7 or higher is appropriate for exploratory research, as it demonstrates adequate internal consistency among scale items measuring the same construct. The overall questionnaire achieved a Cronbach's alpha coefficient of $\alpha = 0.84$, as obtained from the reliability analysis output generated by the Statistical Package for the Social Sciences (SPSS). This value exceeded the recommended threshold of 0.70 and indicated satisfactory internal consistency across all measurement items. Results indicated that the questionnaire items demonstrated acceptable internal consistency. Validity of content was evaluated by examining the relevance, clarity, and alignment of each questionnaire item with the study objectives.

Interview guides first underwent an expert evaluation by two cybersecurity and qualitative research specialists, who assessed them against standards such as question clarity, logical flow, and alignment with study objectives. Two experts were considered sufficient to provide a balanced technical and methodological evaluation while maintaining feasibility. Secondly, three participants from the target audience took part in pilot interviews to evaluate the guide's usability and its ability to elicit thorough, substantive comments. The three participants were sufficient at this stage to identify major issues in question clarity, flow, and response depth before full deployment. Iterative improvements were made to the guide to enhance its clarity, organization, and content based on input from these evaluations and pilot interviews.

Additionally, by providing different viewpoints on the same research issues, triangulation across surveys, interviews, and observations improved validity. To guarantee consistency, observational data were cross-referenced with survey and interview results and confirmed using a systematic

observation checklist. Collectively, these procedures ensured that the data collection tools were appropriate, reliable, and capable of capturing relevant insights on phishing vulnerabilities in digital microfinance institutions.

4.7 Data Analysis

4.7.1 Qualitative Analysis

Following Kogen (2024) six-phase methodology: familiarization, coding, theme creation, review, definition, and report production, a thematic analysis was used to examine qualitative data from interviews. The thematic analysis was a suitable methodology for this study, as it offers flexibility and structure for identifying, analyzing, and interpreting patterns in qualitative data, especially in domains involving human perceptions, experiences, and organizational practices (Braun & Clarke, 2023). Through this technique, responses can be systematically coded and categorized into groups such as cybersecurity procedures, employee attitudes, and organizational vulnerabilities. The thematic analysis was also appropriate for the study, as it helps uncover explicit and implicit meanings in participants' responses, thereby revealing hidden patterns and information regarding phishing risks and cybersecurity practices in microfinance institutions (Rwashdeh et al., 2025). Identifying underlying issues or patterns that might not be directly evident from quantitative data alone was also be aided by thematic analysis.

4.7.2 Quantitative Analysis

Descriptive and inferential statistics were used to assess quantitative survey responses and GRU model performance data: Key variables, such as phishing awareness levels and the frequency of

cybersecurity incidents within institutions, were summarized using measures of central tendency (mean, median) and dispersion (standard deviation).

Regression analysis examined correlations between variables, including the relationship between institutional phishing prevention measures and phishing awareness. Furthermore, Analysis of Variance (ANOVA) tests were employed to determine whether the accuracy of the GRU model's phishing detection varies significantly across phishing attack types or organizational settings.

Before GRU model training and evaluation, feature extraction procedures were applied to the phishing datasets to convert the raw URL, email, and network-traffic data into features suitable for analysis and sequential modelling. For URL data, features such as URL length, domain structure, HTTPS presence, special characters, subdomain complexity, and redirection patterns were extracted. For email data, relevant information was extracted from email headers, sender-domain information, subject lines, message content, embedded links, and attachments, followed by tokenization to obtain sequential representations. For network-traffic data, flow records were sessionized to capture temporal behavior, and features such as connection duration, packet frequency, packet-size characteristics, TCP flags, and byte-distribution patterns were extracted. Derived features such as domain entropy and redirect depth were also generated where applicable.

Following feature extraction, the data were prepared for GRU modelling through normalization of numerical features, one-hot encoding of categorical variables, removal of null values, and sequence padding to ensure consistent input representation. The extracted features were then associated with the corresponding binary labels for phishing and legitimate instances.

4.7.3 Data Analysis Software

Python libraries were employed for data preprocessing and for analyzing the GRU model's performance indicators, and SPSS was used for statistical analysis of survey data. Accuracy, precision, recall, and false positive rates are performance metrics for the GRU model that provide a clear picture of how well it detects phishing attempts in the context of MFIs.

4.7.4 Data Integration

Explanatory sequential analysis combines qualitative and quantitative results within a mixed-methods design. To provide a thorough interpretation of the data, qualitative insights from observations and interviews were used to contextualize quantitative data on the GRU model's performance. This sequential study connects the model's technical effectiveness with useful, context-specific insights from the MFI environment, strengthening the overall conclusions.

4.8 Dataset Strategy

4.8.1 Dataset selection

To support the development and evaluation of a Gated Recurrent Unit (GRU)-based detection framework, this study used three publicly available phishing-related datasets. These datasets cover the main phishing attack surfaces, which include emails, network traffic, and URLs. They were chosen for their structural compatibility with sequential learning, appropriate labeling, and relevance to phishing behaviors commonly observed in digital financial systems.

The first dataset, called the "Phishing Website Dataset," was uploaded on Mendeley Data and was assembled by Verma and Das (2022). It includes 88,647 phishing and legitimate URLs, and characteristics such as domain structure, Hypertext Transfer Protocol Secure (HTTPS) presence,

redirection patterns, special characters, subdomain complexity, and URL length are recorded. When identifying phishing sites that mimic financial institution login portals or transaction pages, these characteristics are essential. The dataset's organized, tokenizable format makes it ideal for GRU-based sequence modeling, and it has proven reliable in phishing detection studies.

The second dataset, the Enron-Spam and Phishing Email Corpus, created by Carnegie Mellon University (2004), contains more than 35,000 labeled email samples, including header information, sender-domain metadata, subject lines, body text, and embedded elements such as links and attachments. The full message content and annotated phishing labels enable modeling of semantic and behavioral cues typical of phishing emails.

The third dataset is the Canadian Institute for Cybersecurity Intrusion Detection System (CICIDS) 2017 Dataset, which was created by the University of New Brunswick's Canadian Institute for Cybersecurity. It includes over 2.8 million individual network records, aggregated into labeled flows representing both benign and attack behaviors, including phishing. The dataset includes features like average packet sizes, connection durations, source and destination internet protocols (IPs), transmission control protocol (TCP) flag values, and byte distribution patterns. It is highly relevant for training sequence-based phishing detection models because it is designed for intrusion detection benchmarking and supports temporal behavior analysis.

These datasets were chosen not just for their individual usefulness but also for their complementarity across the three primary phishing attack surfaces: network traffic, emails, and URLs. To capture both the surface-level and deep sequential characteristics of phishing attempts, the GRU model can learn from syntactic URL patterns, semantic email content, and behavioral network abnormalities thanks to this triangulated dataset approach.

Each dataset is organized using either native sequential data or generated features designed to work with deep learning-based sequence models like GRUs, and each one offers a binary classification objective. Normalization of numerical fields, binary label encoding, sequence padding for temporal characteristics, and one-hot encoding for categorical variables are examples of preprocessing procedures. To ensure compatibility with downstream model training, these operations were carried out using the Scikit-learn and TensorFlow preprocessing libraries.

All datasets were retrieved in April 2025 and are open access, meaning they are licensed for unrestricted scholarly use. To ensure reproducibility and ethical compliance, each was formally cited, with links to the sources included in the appendix.

4.8.2 Dataset Justification and Relevance

These datasets were selected for their methodological adequacy and thematic alignment with phishing risks observed in Nairobi's microfinance institutions (MFIs). According to local reports and case studies, phishing occurrences in the Kenyan banking industry frequently entail fake portals, email impersonation, and unauthorized access—threats that are structurally and behaviorally duplicated in the selected datasets (Muriuki & Wamuyu, 2022).

Nevertheless, because of their global origin, these datasets may not accurately reflect the linguistic, cultural, or technical aspects of phishing efforts targeting users at MFIs in Nairobi. To overcome this constraint, a hybrid data approach was utilized. The university's ethics council approved the acquisition of localized phishing data, including incident logs, staff-reported attacks, and email templates, from participating MFIs in accordance with Kenya's Data Protection Act (2019). Region-specific attack scenarios were simulated through structured interviews and phishing case reconstructions when direct access to institutional logs is unavailable.

Using this local data, domain adaptation techniques were used to refine the pretrained GRU model. To reduce false positives in delicate operating situations, these methods involved reweighting samples, applying specific token embeddings for local context, and modifying classification criteria.

This approach guarantees the contextual validity of applying results to the Kenyan microfinance ecosystem as well as the technical integrity of using validated datasets. By combining locally generated data with benchmark datasets, the model was more accurate in identifying phishing attempts specific to the Nairobi digital financial industry.

4.9 Ethical Considerations

4.9.1 Informed Consent

Before data collection started, all participants received a thorough explanation of the study's objectives, data processing procedures, and their rights as participants. Participants were made aware of their right to withdraw from the study at any time, and written informed consent was sought.

4.9.2 Confidentiality and Data Privacy

Confidentiality of data was carefully upheld. To prevent participant identification, personal identifiers were eliminated from survey and interview responses, and institutional data were anonymized. Only authorized researchers had restricted access to the data, which was kept on protected digital storage devices.

4.9.3 Ethical Approval

The National Commission for Science, Technology, and Innovation (NACOSTI) and the ethics review board of the associated university were consulted for ethical approval. In particular, regarding participant rights, data privacy, and cybersecurity, this study abided by institutional and national research ethics guidelines.

4.9.4 Addressing Ethical Challenges in Real-Time Model Testing

Ethical issues could arise during the GRU model's experimental testing phase if it unintentionally flags legitimate emails as phishing. Testing was carried out under strict supervision in a controlled setting to reduce these risks, and institutional IT staff reviewed any highlighted findings to minimize interference and prevent false positives.

4.10 Summary

This chapter describes the study approach used to examine phishing vulnerabilities among microfinance institutions (MFIs) in Nairobi, Kenya. The combination of qualitative and quantitative methodologies in the mixed-methods approach guarantees a thorough examination of the technical and contextual aspects of the study subject. The qualitative elements offer a profound understanding of organizational weaknesses, while the quantitative techniques thoroughly assess the GRU neural network model's efficacy.

The sampling techniques, data collection methods, and research methodologies were specifically selected to achieve the study's intended objectives. To ensure the study's legitimacy and integrity, participant approval and data confidentiality were considered. To ensure that the results are both

actionable and scientifically sound, the chapter provides a strong foundation for the research's later phases, including data collection, analysis, and AI model development.

CHAPTER 5

RESULTS AND DISCUSSION

5.1 Introduction

This chapter outlines the primary findings and analyses related to the study's four objectives. These objectives included mapping cybersecurity vulnerabilities and phishing threats in digital microfinance institutions (MFIs) that can be addressed through AI-driven solutions; assessing the effectiveness and applicability of both global and local cybersecurity and AI governance frameworks in mitigating phishing threats within digital microfinance institutions; developing a gated recurrent unit (GRU)-based phishing mitigation model based on the MFI settings; and validating this model on empirical data.

5.2 Cybersecurity Vulnerabilities and Phishing Threats in Digital Microfinance Institutions

The findings confirm that phishing remains the most persistent cybersecurity threat in financial services, underscoring the need for an AI-driven response that integrates technical innovation and regulatory awareness. To address that requirement, the chapter combines vulnerability mapping, governance analysis, and GRU model development, aligning with recent scholarly calls for integrated approaches that combine empirical evidence, policy critique, and advanced machine learning (Shahid et al., 2022; Mahmud et al., 2024). By positioning phishing as the central vulnerability and GRU neural networks as a practical, context-sensitive solution, the chapter contributes to both immediate risk management strategies for MFIs and the broader scholarly debate on AI adoption in financial cybersecurity.

5.2.1 Data Sources and Analytical Approach

The analysis presented in this chapter is based on several data sources and thus enables triangulation, which adds to the credibility and depth of the findings. Primary data were collected through surveys, interviews, and phishing simulations at Kenyan microfinance institutions. These approaches facilitated the acquisition of user-level vulnerability and institutional readiness, as well as sector-specific vulnerability patterns. The use of phishing simulations was especially important, as it offers an experimental evaluation of exposure to real risks- a drawback of self-reported surveys (Purkait et al., 2021). The sources of secondary data were international, regional, and national cybersecurity frameworks, such as the NIST Cybersecurity Framework, the EU Artificial Intelligence Act, the Model AI Governance Framework in Singapore and in Kenya, the Data Protection Act (2019), the Computer Misuse and Cybercrimes Act (2018), and the National Cybersecurity Strategy (2022-2025). The analysis of these frameworks helped assess their suitability for integrating artificial intelligence to mitigate phishing in the financial sector. Lastly, technical data were acquired by developing a phishing-mitigation model based on a gated recurrent unit (GRU).

GRUs were chosen for their proven ability to process sequential data, such as phishing messages, with high computational efficiency relative to long short-term memory networks (LSTMs) while maintaining high accuracy (Huang et al., 2020). To ensure that the selection of the GRUs was evidence-based, the performance of the selected GRUs was benchmarked against alternative deep learning architectures, including LSTM, bidirectional LSTM (BiLSTM), and convolutional neural networks (CNNs).

By bringing together behavioral, policy, and technical data, this layered design takes a socio-technical systems approach to understanding cybersecurity, recognizing that technical solutions

are most effective when integrated into organizational practices and governance structures (Hoong et al., 2024)

5.2.2 Rationale for Focusing on Phishing in MFIs

Phishing lies at the center of the current study because of its widespread and global nature and because of its relatively devastating consequences to financial institutions located in low- and middle-income countries. According to threat-intelligence reports, phishing continues to be the most common vector used to carry out most of the cyber-attacks against the financial sector (Verizon, 2023; Symantec, 2022). Within the Kenyan context, phishing has been directly linked to mobile-money fraud, unauthorized system access, and business email compromise scams, which are often targeted against micro-finance institutions (CA, 2022). These vulnerabilities are compounded by the sector's dependence on digital communication channels (especially email and short message service (SMS)), which are currently the most exploited vectors for phishing attacks.

Microfinance institutions face additional structural vulnerabilities. Limited cybersecurity budgets, coupled with high staff exposure to customer-facing communications, increase susceptibility. Moreover, while Kenya has established a growing body of legislation, including the Data Protection Act (2019) and the Computer Misuse and Cybercrimes Act (2018), these frameworks do not yet provide explicit guidance on AI-driven defenses. The absence of a regulatory mechanism for algorithmic transparency, model certification, or accountability in AI adoption further exacerbates institutional risk. For these reasons, phishing is not treated here as one among many vulnerabilities but as the primary threat vector through which AI-driven detection strategies can be meaningfully developed and evaluated.

5.3 Objective 1 – mapping cybersecurity vulnerabilities facing digital MFIs that can be addressed through AI-driven solutions

5.3.1 MFI Cyber Threat Landscape

The results of this study established phishing as the most significant cybersecurity vulnerability facing microfinance institutions (MFIs) in Kenya. In controlled phishing simulations, 36% of staff clicked on malicious links, confirming high levels of employee susceptibility to social engineering. Departmental analysis revealed that Finance staff were the most vulnerable (78.1%), followed by Customer Service staff (71.9%) and IT personnel (69.3%). These findings align with global evidence, such as the Verizon Data Breach Investigations Report (2023), which attributes 74% of breaches to human factors, primarily phishing. The high exposure of IT personnel — often assumed to be more cyber-aware — underscores the persistence of human error even among technically skilled staff.

The survey data was further mapped into the forms of phishing attacks. Malicious attachments or links in emails were the most frequently reported (61.2%), followed by email spoofing or impersonation (51.0%). SMS or WhatsApp phishing (smishing) was cited by 44.9% of respondents, while 26.5% reported fake websites or login portals, and 18.4% highlighted social media impersonation of staff or institutions. This breakdown confirms that phishing is not confined to email but extends into mobile and social channels, which are increasingly integral to the MFI digital ecosystem.

Beyond phishing, respondents identified a spectrum of other vulnerabilities: mobile money fraud, credential theft and account compromise, ransomware, insider threats, and data breaches from unauthorized access. Importantly, many of these threats are interlinked with phishing, which serves as a gateway vulnerability enabling further compromise. For instance, phishing-related credential theft often escalates into account takeovers (ATO) or fraudulent withdrawals. Similarly, rogue loan applications exploit trust in MFI brands, using phishing campaigns as the entry point.

Targeted staff roles were also consistent across both the simulations and the survey. Finance and IT departments were the most frequently targeted, while customer-facing staff, particularly those in front desk and support roles, were also reported as highly exposed. This reflects attackers' strategic focus on individuals with access to sensitive transactions or those frequently interacting with clients.

From a socio-technical perspective, these vulnerabilities reflect weaknesses across human, technical, and institutional subsystems (Hoong et al., 2024). Weak awareness programs with only 22 percent of staff receiving quarterly training are compounded by technical gaps, such as the absence of multi-factor authentication in 68% of MFIs, and institutional shortcomings, including 61.4% lacking phishing-specific incident response plans. Collectively, these weaknesses create fertile ground for social engineering attacks to succeed.

In mapping these vulnerabilities against their suitability for AI-driven mitigation, it is evident that phishing, smishing, and credential harvesting are the most AI-ready risks. They produce large volumes of sequential text, communication, and behavioral data — email logs, SMS patterns, login records — that can be effectively modeled by recurrent neural architectures such as GRUs. Mobile money fraud and account takeover can also be partially addressed through AI, particularly when linked to anomalous transaction monitoring. In contrast, threats such as ransomware and insider

abuse are not as easily addressed by deep learning alone. They usually need hybrid defenses that combine AI and rule-based surveillance with robust organizational protection. Altogether, the results confirm the fact that phishing is the most urgent and AI-resolvable threat to MFIs. Generally, it is a point of entry to a bigger ecosystem of risks. Since phishing takes numerous forms, including malicious emails, SMS smishing, and even social engineering, it generates systematic data that can be effectively learned through adaptive neural networks. Other cyber risks, like mobile money fraud and credential theft, can be partially mitigated with AI; however, risks such as ransomware and insider abuse highlight the limits of AI-only solutions. This mapping provides a fair representation of the vulnerability threat landscape and the potential of GRU-based models, thereby preparing the next section on AI-based mitigation strategies.

5.3.2 AI-Driven Mitigation of Phishing and Related Threats

As highlighted in Section 4.2.1, the vulnerabilities signify that phishing is often a precursor to more sophisticated attacks, including account takeover, credentials theft, and mobile money fraud. These risks vary in scale and nature, but they have one thing in common: they are based on sequential patterns in communication and user behavior that are difficult to detect with rule-based, static defenses. This highlights the need for adaptive security approaches that can scale with evolving attack vectors, rather than security mechanisms that remain fixed, such as filters or firewalls.

This adaptive capability comes from artificial intelligence (AI), especially from recurrent deep learning architectures such as Gated Recurrent Units (GRUs). By analyzing data chronologically - whether it is the level of urgency of a phishing email, the character-level structure of a fraudulent SMS, or the irregular timing of login attempts, GRUs can uncover hidden patterns that signal malicious intent. Unlike signature-based filters that rely on prior knowledge, GRUs generalize to

detect novel, previously unseen attacks, making them particularly well-suited to the dynamic threat landscape faced by MFIs (Mahmud et al., 2024).

In practice, the vulnerabilities identified earlier map differently onto AI readiness. Phishing emails and spoofing attempts, reported by 61.2% and 51% of respondents, respectively, are highly AI-ready because they generate structured textual data. GRUs can detect shifts in urgency in business email compromise (BEC) scams, recognize suspicious URL embeddings, and flag repeated, near-identical campaign bursts that overwhelm human defenders. Similarly, smishing, experienced by 44.9% of respondents, is well-suited to GRU modeling: character-level analysis enables detection of fraudulent mobile money alerts disguised as M-Pesa or loan repayment reminders.

Credential harvesting through fake portals (26.5% of reports) also lends itself to GRU mitigation. Here, sequential models analyze URL structures, click patterns, and request logs to identify subtle anomalies that distinguish fraudulent portals from legitimate ones. These capabilities extend naturally to account takeover (ATO) attempts, where GRUs can track deviations in login sequences — such as unusual time-of-day access, repeated password resets, or access from unrecognized devices. While ATO and mobile money fraud fall into a “medium readiness” category, GRUs enhance anomaly detection and provide early warning signals that, when combined with rule-based transaction controls, reduce the likelihood of undetected breaches.

By contrast, ransomware and insider threats are less well-suited to AI, as they rely more on system compromise or privilege abuse than on sequential communication data. Here, GRUs contribute indirectly by flagging anomalies in user behavior, but effective defense requires complementary endpoint monitoring, access governance, and institutional controls. This distinction underscores that AI is not a universal solution; rather, it must be strategically targeted to vulnerabilities that generate patterns suitable for sequential learning.

Nevertheless, GRU-driven defenses are not without challenges. High false-positive rates risk overwhelming already resource-constrained IT teams, and the “black-box” nature of GRU predictions can undermine staff trust. To address this, explainable AI (XAI) tools such as SHapley Additive exPlanations (SHAP) and Local Interpretable Model-agnostic Explanations (LIME) have been integrated into phishing detection models to generate simplified justifications, for example, “suspicious sender history” or “unusual link structure,” which improve both adoption and compliance (Ribeiro et al., 2022; Lundberg et al., 2023). Even with these advances, attackers are adapting adversarial phishing designed to bypass AI classifiers, which is already emerging, illustrating the ongoing arms race between defenders and adversaries (Fang et al., 2022).

In summary, GRUs provide the strongest defense against the vulnerabilities most prevalent in Kenyan MFIs' phishing, smishing, and credential harvesting while also enhancing anomaly detection for account takeover and mobile fraud. They are less effective against ransomware or insider abuse, highlighting the need for hybrid socio-technical defenses. Importantly, their successful deployment requires integration into broader institutional ecosystems of training, response planning, and regulatory oversight. This sets the stage for the next section, which evaluates whether Kenya's current governance and policy frameworks are prepared to support the responsible adoption of AI-based defenses in microfinance institutions.

5.4 Objective 2 – Cybersecurity and AI Policy Frameworks

This section examines governance structures at global, regional, and national levels, with a focus on how effectively they address cybersecurity, artificial intelligence (AI), and phishing threats in digital microfinance institutions (MFIs). For MFIs, many of which operate with limited resources,

governance is not an abstract concern; it determines whether advanced tools such as the GRU phishing mitigation framework can be adopted safely and responsibly.

This is analyzed from three perspectives. The first one is the responsive regulation, which relates to a more adaptive regulation in accordance with the degree of risk, balancing the need to innovate with relevant safeguards (Johnson, 2023; Hadfield, 2023). The second one is institutional theory, which considers the process of adoption or non-adoption of global norms in organizations. Recent research shows that, with AI, institutions regularly imitate international norms in order to gain legitimacy, whereas smaller actors, including MFIs, may implement them symbolically due to limited resources (EURAM, 2025; ICIS, 2023). The third approach is risk-based and responsible AI governance, but this approach acknowledges that technologies vary in the degree of oversight they need. More accountability and transparency are required in high-risk finance-related applications, including phishing detection, and less regulation can be imposed on low-risk systems (Papagiannidis, 2025; Nature, 2024). Combined, these views indicate that phishing is not merely a technical problem. It is a challenging and high-risk AI problem that needs sufficiently robust governance structures to protect MFIs and their clientele, while remaining flexible enough to encourage innovation.

5.4.1 Global Cybersecurity Frameworks Overview

The ISO/IEC 27001, NIST Cybersecurity Framework (CSF), Control Objectives for Information and Related Technologies (COBIT), and the General Data Protection Regulation (GDPR) are globally accepted benchmarks for information security. However, to most microfinance institutions (MFIs) in Kenya, these standards are more of a dream than a reality. Survey findings indicated that merely 19 per cent of MFIs stated full compliance with any of the global framework with the majority of them citing the high costs, lack of technical know-how as well as lack of

localized guidelines as the main impediments. This, in practice, implies that numerous MFIs depend on their own internal policies (or improvised policies), thus creating an unequal and non-uniform safeguard within the sector.

Although these conventional structures are strong in overall security procedures, they remain silent in artificial intelligence (AI). In comparison, recent laws like the EU Artificial Intelligence Act and the Singapore MAS Principles specifically address how AI is meant to be used in the financial services sector. Such frameworks introduce explainability, fairness, and accountability requirements that are particularly applicable to phishing mitigation. Other jurisdictions have also released principle-based guidance on AI, notably India (NITI AIZ Strategy), Australia (AI Ethics Framework), and China (AI Security Draft), but with less comprehensive approaches to sector-specific risks such as phishing.

Importantly, the survey findings indicated that none of the Kenyan MFIs have implemented an AI-specific governance model, which reaffirms the lack of alignment between the new international standards and the local frameworks. Table 4.1 provides a summary of global cybersecurity frameworks, showing their relevance and AI readiness

Table 5.1 Global Cybersecurity and AI Governance Frameworks

Framework	**Jurisdiction**	**Relevance to Phishing**	**AI Readiness**
NIST CSF	USA	Strong on identity/access controls	No explicit AI guidance
ISO/IEC 27001	Global	Comprehensive ISMS	No AI provisions

COBIT	Global	IT governance and control	No AI provisions
GDPR	EU	Data protection, consent, privacy	Only indirect references to AI
EU AI Act	EU	Classifies phishing detection as high-risk	Yes
MAS AI Principles	Singapore	Finance-sector AI governance	Yes
NITI AI Strategy	India	Promotes ethical, inclusive AI	Partial
AI Ethics Framework	Australia	Fairness, explainability principles	Partial
AI Security Draft	China	Centralized, risk-driven	Limited detail

Source: Author's own compilation

The findings suggest that while MFIs are aware of global best practices, adoption is selective and often partial. Consistent with institutional theory, organizations tend to adopt only those elements that are feasible within their financial and technical constraints, leading to wide variations in protection against phishing threats.

5.4.2 Legal and Policy Gaps in Kenya's Cybersecurity Framework

Kenya's legal foundation for cybersecurity is built on the Data Protection Act (2019), the Computer Misuse and Cybercrimes Act (CMCA, 2018), the Kenya Information and Communications Act

(1998), and the National Cybersecurity Strategy (2014). When they were introduced, these laws placed Kenya among regional leaders in digital security. However, they were developed before the rapid rise of AI-driven threats and therefore do not address how technologies such as phishing detection models should be governed. Table 4.2 provides an overview of Kenya's current cybersecurity legal and policy frameworks, assessing their effectiveness in addressing emerging challenges posed by artificial intelligence (AI), particularly in phishing prevention. The table indicates that although Kenya has built a strong foundation for cybersecurity and data protection, there are significant gaps in its frameworks regarding the regulation of AI-driven threats and solutions.

Table 5.2 Kenyan Cybersecurity Legal Frameworks and AI Gaps

Law/Policy	**Relevance**	**Gap for AI in Phishing Mitigation**
Data Protection Act (2019)	GDPR-aligned, regulates personal data	No mandates on AI transparency or bias audits
The Computer Misuse and Cybercrimes Act (CMCA) (2018) and CMCA Amendment 2025	Criminalizes phishing and cybercrime	Silent on AI explainability, auditing, or sandboxing
National Cybersecurity Strategy (2014).	Broad policy on cyber resilience	Outdated, pre-AI, lacks provisions for adaptive models
KICA (1998)	ICT infrastructure regulation	Legacy law, no AI integration

Source: Author (2025).

Survey results mirror these gaps. While the CMCA makes phishing a criminal offense, 45% of respondents were unaware of their obligations under the Data Protection Act, and 31.6% of MFIs reported having no incident response plan. Just as importantly, none of the institutions surveyed had adopted an AI-ready governance framework, highlighting the absence of clear national guidance on how AI should be introduced and overseen in financial services.

This gap is also evident in official reviews. The Communications Authority (CA, 2022) has pointed to limited funding and a shortage of skilled auditors capable of assessing AI tools. Similarly, the Central Bank of Kenya (CBK, 2023) issued updated cybersecurity guidelines for financial institutions, but these did not mention AI-driven defenses. As a result, adoption remains unregulated. Research further shows that MFIs, unlike larger banks, often rely on off-the-shelf digital tools without the resources to independently check their accuracy or fairness (Mhlanga, 2024).

Collectively, the evidence shows that while Kenya's laws prohibit phishing, they stop short of regulating the AI systems that could be used to prevent it. MFIs, therefore, operate in a space where cybercrime is illegal, but AI-based defenses lack any formal oversight.

5.4.2.1 Policy Misalignments in MFI Governance

Empirical survey data from Kenyan MFIs reinforces these weaknesses. Despite legal reforms, 31.6% of surveyed MFIs reported having no incident response policy, while 45% were unaware of their obligations under the Data Protection Act. Critically, none reported adopting AI-specific governance frameworks. As demonstrated in Table 4.3, many of Kenya's MFIs are not well prepared to address contemporary cybersecurity issues. Over one-third of organizations lack an incident response plan; almost half of MFIs are unaware of their duties under the Data Protection Act; and none have developed guidelines for governing artificial intelligence technologies.

Table 5.3 Policy Misalignments in MFI Cybersecurity Governance

Issue	**Finding (Kenya MFIs)**	**Implication for AI & Phishing**
Incident Response (IR) Gaps	31.6% lack IR policies	Limited resilience to phishing
Regulatory Awareness	45% unaware of Data Protection Act duties	Risk of non-compliance, liability
AI Governance	0% adopted AI frameworks	Opaque, unaccountable AI use

Source: Author (2025).

These results illustrate what Johnson (2021) calls the implementation gap: laws exist, but enforcement and compliance remain weak. From an institutional theory perspective, MFIs face little coercive pressure (regulators rarely audit or sanction) and weak normative pressure (few industry-wide AI standards). Consequently, they may adopt GRU-based phishing detection tools in an unregulated space, exposing them to risks of bias, accountability failures, and consumer mistrust.

The lack of AI-ready regulation leaves MFIs "caught in the middle": on one hand, they must comply with static laws, while on the other, they face growing pressure to adopt adaptive AI defenses to keep pace with evolving phishing threats. Without substantive regulatory reform, the unregulated use of these technologies risks causing unintended harms, including the mistaken flagging of legitimate transactions or the processing of sensitive customer data without proper oversight.

On the other hand, this regulatory gap provides an innovative platform. Kenya, already touted as a fintech hub, has the potential to cement its leadership status by enacting future-focused policies,

including AI transparency requirements, regulatory sandboxes for deploying cybersecurity apps, and sector-wide standards implemented in collaboration with microfinance institutions. From the perspective of responsive regulation (Ayres & Braithwaite, 2023), the law must go beyond criminalizing phishing to include prescriptive guidance on how AI can be used as a defensive mechanism. Such improvements to this legislation would also enhance compliance and consumer confidence in the digital financial ecosystem. In accordance with these empirical findings, the next section presents the proposed changes to the Computer Misuse and Cybercrimes Act of 2018 and its 2025 amendment, with the clear intention of further enhancing AI preparedness.

5.4.3 Proposed Amendments to the CMCA (2018)

Empirical survey data on Kenyan micro-finance institutions demonstrate blatant gaps in the interpretation and application of the existing cybersecurity regulations. Approximately a third (31.6%) of the surveyed MFIs reported not having adopted an incident-response policy, thereby lacking a formal mechanism to address phishing attacks. Meanwhile, awareness of the Data Protection Act of 2019 was also unsatisfactory, as 45 per cent of respondents stated they were unaware of what the law required. Even worse, none of the surveyed MFIs had an AI-ready governance structure. Collectively, these findings highlight a critical gap: the CMCA of 2018 makes phishing illegal; however, it does not prescribe how AI systems designed to mitigate such threats should be regulated or supported by institutions. Global regulatory frameworks provide essential comparative benchmarks. One possible amendment is to establish regulatory sandboxes, which was monitored and coordinated by the Central Bank of Kenya (CBK) and the National Kenya Computer Incident Response Team – Coordination Centre (KE-CIRT/CC). Such sandboxes would help MFIs to test AI-enhanced phishing mitigation technologies in a controlled

environment. This would respond to the survey finding that almost a third of the institutions had no incident response plan whatsoever.

The second area in which reform is needed concerns the accountability of AI decisions. Whereas phishing is already a criminal offense, the liability of the error committed by the AI system is not specified anywhere in the CMCA, such as the failure to block legitimate transactions or the inability to detect phishing. This gap gains even more significance in light of the findings of the survey that show that almost fifty percent of MFIs are poorly aware of their current compliance requirements.

The survey also revealed quite vividly the absence of independent verification practices, with 0 per cent of MFIs having any AI governance practices deployed. That requirement increases the necessity of establishing a national AI certification system, which would be managed by the Kenya Bureau of Standards (KEBS) and KE-CIRT/CC, and would offer a systematic audit of robustness, fairness, and resistance to manipulation.

Another remarkable weakness is the lack of transparency. Even though the Data Protection Act regulates personal data, it does not require AI vendors to disclose how their systems work. The fact that 45% of MFIs were unaware of their data protection responsibilities suggests that, without explicit mandates, many institutions will continue to operate with limited oversight of AI tools.

The lack of sector-specific guidance is equally evident. With only 19% of MFIs reporting compliance with any global cybersecurity standard, most continue to rely on improvised or internal policies that vary widely in scope and quality. Tailored guidelines from the CBK and Capital Markets Authority (CMA) could therefore provide a baseline for smaller institutions.

Finally, introducing a risk classification framework would enable regulators to distinguish between low-risk applications, such as customer chatbots, and high-risk systems, such as phishing detection. The survey findings revealed that none of the MFIs had adopted AI-ready frameworks, reinforcing the conclusion that the sector currently lacks a structured way to differentiate risks.

Collectively, these findings illustrate the areas where the CMCA (2018) could be strengthened. Survey results—showing that 31.6% of MFIs lacked incident response policies, 45% were unaware of their data protection obligations, and none had adopted AI frameworks—illustrate just how wide the gap is between Kenya's current legal provisions and the practical needs of MFIs. These gaps form the regulatory backdrop for the next section, which turns to the technical innovations explored in this study, namely the development and validation of a GRU-based phishing detection model. Table 4.4 below consolidates these proposed reforms, responsible entities, and potential risks.

Table 5.4 Proposed Amendments to CMCA (2018) for AI Readiness

Proposal	**Description**	**Responsible Entity**	**Potential Risk**	**Survey Findings**
Regulatory Sandboxes	Allow supervised AI pilots (e.g., MAS model)	CBK, KE-CIRT/CC	Risk of weak oversight being exploited	31.6% of MFIs reported no incident response policy, showing the need for structured

				testing environments
AI Decision Accountability	Define liability & appeals for AI outcomes	Parliament, Judiciary	Litigation burden, regulatory capture	45% of MFIs are unaware of DPA obligations, highlighting weak awareness of compliance responsibilities
National AI Certification	Independent audits for fairness, robustness	KEBS, KE-CIRT/CC	Cost burden on smaller MFIs	0% of MFIs reported AI-ready frameworks, indicating no current validation mechanisms
Algorithmic Transparency	Require model documentation, bias reports	Data Commissioner	Risk of exposing proprietary models	45% unaware of DPA duties, suggesting lack of awareness around data

				handling and AI explainability
Sector-Specific AI Guidelines	Tailored for fintech and MFIs	CBK, CMA	Compliance burden for smaller MFIs	Only 19% of MFIs comply with global standards, leaving the majority dependent on improvised policies
AI Risk Classification	Tiered oversight (high, medium, low-risk AI)	National AI Taskforce	Difficulty in categorizing hybrid systems	0% of MFIs had adopted AI-ready frameworks, showing absence of structured risk differentiation

Source: Author (2025).

5.4.3.1 Implications and Implementation Challenges

The survey evidence suggests that implementing AI-ready reforms will not be straightforward for MFIs in Kenya. None of the institutions surveyed (0%) had adopted an AI governance framework, 31.6% reported having no incident response policy, and 45% of respondents indicated that they

were unaware of their obligations under the Data Protection Act (2019). These figures point to an environment in which phishing attacks disproportionately target institutions yet lack the capacity—financial, technical, or regulatory awareness—to validate and monitor advanced AI tools. In such a context, measures such as national certification or sector-specific guidelines would provide much-needed accountability. However, they would also create compliance demands that many MFIs are not currently prepared to meet.

The challenges are not confined to MFIs alone. Regulators tasked with overseeing the sector also face significant limitations. The Communications Authority of Kenya (CA, 2022) has reported persistent funding shortages and a shortage of skilled auditors capable of assessing AI-driven systems. Similarly, the Central Bank of Kenya (CBK, 2023) has published cybersecurity guidelines requiring financial institutions to establish risk management frameworks. Yet these guidelines do not address AI-driven defenses, leaving adoption largely unregulated. Comparative evidence from the European Union highlights that even in more resource-endowed settings, defining liability for AI decisions has faced industry resistance due to fears of litigation and compliance costs (Ebers, 2023).

Collectively, these findings highlight a dual challenge: MFIs lack the internal resources to comply with AI-ready regulations, and national regulators lack the capacity to enforce them. This combination explains why, despite the urgency of the threats, reforms to integrate AI governance into Kenya's legal framework have been slow to materialize. It also provides the institutional backdrop for the next section, which shifts attention from the policy environment to the technical development of a GRU-based phishing detection model designed to address the vulnerabilities identified in the survey.

5.5 Objective 3 – GRU-Based Phishing Detection Model

The gaps in the regulatory and institutional frameworks identified in the previous section necessitate the urgent introduction of technical solutions that can operate efficiently in resource-limited microfinance institutions (MFIs). This section introduces the GRU-based phishing mitigation model developed in this research, which addresses the third research objective: designing a phishing mitigation system responsive to the realities of digital MFIs. The selection of Gated Recurrent Units (GRUs) can be justified by the fact that the threat environment confronting MFIs is highly context-specific, with phishing attacks frequently taking the form of smishing targeting mobile users, spear-phishing e-mails targeting finance teams, and business e-mail compromise scams involving executive impersonation. To mitigate these risks, it is necessary to have a detection strategy that is not only accurate but also lightweight, scalable, and adaptable to the MFI environment, which is characterized by limited technical resources and a lack of trained personnel (Mahmud et al., 2024).

5.5.1 GRU Justification

Gated Recurrent Units (GRUs) are the most appropriate models to consider for phishing identification because they can learn sequential dependencies in text data while remaining computationally efficient. GRUs have two gating mechanisms, update and reset gates, compared to Long Short-Term Memory (LSTM) networks, which have three gating mechanisms, thereby reducing the number of parameters and enabling faster convergence in training (Cho et al., 2014). This computing efficiency is helpful in resource-constrained environments, such as MFIs, where computational resources and storage may be restricted.

Recent studies indicate that GRUs are more competitive in identifying phishing attacks than other recurrent neural networks. As an example, Birthriya et al. (2024) demonstrated that hybrid GRU-

CNN models achieved higher phishing-site detection rates, and Mahmud et al. (2024) established that GRU-enhanced classifiers were particularly well-suited to detecting smishing attacks due to their ability to identify anomalies in short text chains. GRUs, therefore, provide a good trade-off between accuracy, efficiency, and deployability in MFIs.

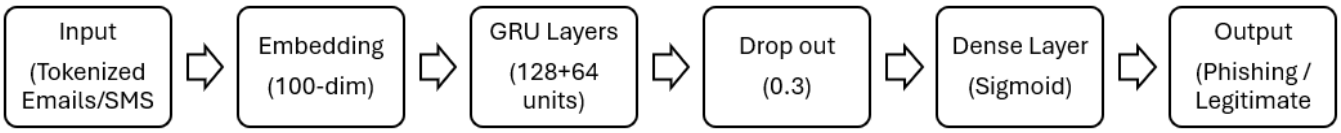
5.5.2 Dataset Preparation

The GRU model was trained and evaluated using the three publicly available datasets described in Section 3.8.1, namely the Mendeley Phishing Website Dataset, the Enron-Spam and Phishing Email Corpus, and the CICIDS 2017 Dataset. These datasets were selected because they collectively represent the primary phishing attack surfaces of URLs, emails, and network traffic, thereby providing comprehensive coverage of phishing behaviors relevant to digital microfinance institutions.

Before model training, each dataset underwent preprocessing to ensure compatibility with the GRU architecture while preserving the sequential characteristics required for deep learning. The main data preparation stages were as follows:

- Tokenization: Splitting text into tokens (words or sub-words).
- Vectorization and Embedding: Converting tokens into dense vectors using a 100-dimensional embedding layer.
- Normalization: Text inputs were normalized through lowercasing, stop word removal, and lemmatization to ensure consistency and reduce noise.
- Padding: Ensuring uniform sequence lengths for compatibility with GRU input.
- Data Partitioning: A stratified 70:20:10 split into training, validation, and test sets to preserve phishing/non-phishing ratios.

Because the three datasets represent complementary phishing attack vectors, the combined dataset enabled the GRU model to learn syntactic URL patterns, semantic characteristics of phishing emails, and temporal network traffic behavior within a unified learning framework. This improved the model's ability to detect phishing attacks across multiple digital communication channels while enhancing the generalizability of the developed framework. Figure 4.1 below illustrates the overall process of data flow of the GRU phishing detection model pipeline



Source: Author (2025).

Figure 5.1 GRU phishing detection model pipeline

5.5.3 Model Architecture

The phishing mitigation model, based on GRU, was created using the deep learning framework TensorFlow, which integrates Keras as its high-level application programming interface (API). The architecture was developed to extract both the semantic content and the temporal patterns embedded in phishing messages, enabling accurate recognition of malicious emails and SMS messages.

The pipeline model started with an input layer that took tokenized text sequences; the sequences were padded to a fixed maximum length to ensure uniformity. These sequences were then fed through an embedding layer, in which words were represented as 100-dimensional dense vectors, thereby preserving semantic relationships in the text.

The architecture was based on two stacked GRU layers with 128 and 64 units, respectively, to model sequential dependencies. A dropout value of 0.3 was used to separate them to reduce

overfitting. These features were then fed into a dense output layer with a sigmoid activation function to classify individual messages as phishing or legitimate.

The Adam optimizer was used to train the model with a learning rate of 0.001 and binary cross-entropy as the loss. Training employed an early stopping strategy with a patience value of three epochs. This means that model training was automatically terminated if the validation loss failed to improve for three consecutive epochs, thereby preventing unnecessary training and reducing the risk of overfitting. Under this stopping criterion, the model converged after 23 training epochs, which represents the actual number of epochs completed before training was terminated. Table 4.5 provides a summary of the GRU model architecture

Table 5.5 GRU Model Architecture

Component	**Details**
Input Layer	Tokenized email/SMS sequences
Embedding Layer	100-dimensional word embeddings
GRU Layers	Two stacked GRU layers (128, 64 units)
Activation	ReLU → Sigmoid for classification
Dropout	0.3 (to reduce overfitting)
Dense Layer	Fully connected with sigmoid activation
Loss Function	Binary Cross-Entropy
Optimizer	Adam (LR = 0.001)
Epochs	Training converged after 23 epochs using early stopping (patience = 3 epochs)

Source: Author (2025).

The code snippet below illustrates the architecture of the GRU model. It illustrates how to build a GRU-based neural network using TensorFlow's Keras Sequential model for text data. It begins with an embedding layer, which turns words into numerical vectors so the model can better understand their meanings and relationships. These vectors are then passed through two GRU layers. The first GRU layer, with 128 units, processes the entire sequence and learns patterns by tracking how words relate to one another over time. A dropout layer is added next to help prevent overfitting by randomly switching off some neurons during training, making the model more robust to new data. The second GRU layer, with 64 units, further processes the information and produces a more compact representation of the sequence. Overall, this setup is well-suited for text-related tasks because it can effectively capture both word meaning and word order.

Figure 4.2 provides an illustrative implementation of the proposed GRU architecture. The code demonstrates the sequential arrangement of the embedding layer, stacked GRU layers, dropout regularization, and the sigmoid output layer used for binary phishing classification. It is presented to illustrate the implementation logic of the proposed model rather than the complete training pipeline. The full implementation, including data preprocessing, hyperparameter tuning, model training, validation, and testing procedures, is provided in Appendix 9.

```
 from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Embedding, GRU, Dropout, Dense


model = Sequential()

model.add(Embedding(input_dim=vocab_size, output_dim=100, input_length=max_len))

model.add(GRU(128, return_sequences=True))

model.add(Dropout(0.3))
```

```
model.add(GRU(64))
model.add(Dense(1, activation='sigmoid'))


model.compile(optimizer='adam', loss='binary_crossentropy', metrics=['accuracy'])
```

Figure 5.2 Illustrative code snippet

5.5.4 Model Performance and Comparison

GRU model performance was compared with a range of other deep learning models, including LSTM, BiLSTM, CNN, and a standard RNN. The main metrics compared were accuracy, recall, F1-score, Receiver Operating Characteristic – Area Under Curve (ROC-AUC), general deployment suitability, and computational latency in MFIs.

Accuracy measures the proportion of all predictions that were correctly classified. Precision measures the proportion of instances predicted as phishing that were actually phishing, whereas recall measures the proportion of actual phishing instances that were correctly detected. F1-score is the harmonic mean of precision and recall and therefore provides a balanced measure of the two. ROC-AUC measures the model's ability to discriminate between phishing and legitimate instances across different classification thresholds, with values closer to 1 indicating stronger discriminative performance (Almiani et al., 2023; Islam et al., 2022).

The dataset was stratified into 70:20:10 training, validation, and test sets to maintain the phishing-to-non-phishing ratio. To ensure a consistent and objective comparison, all models were evaluated using the same data partitions, preprocessing pipeline, optimizer (Adam), loss function (binary cross-entropy), batch size, early stopping strategy, and evaluation metrics. Hyperparameter

settings specific to the baseline models were adopted from commonly reported configurations in the literature rather than being independently optimized for this study. Consequently, the comparison was intended to provide a consistent performance benchmark under a common experimental protocol rather than an exhaustive hyperparameter optimization of each competing architecture. Table 4.6 summarizes the comparative performance of the evaluated models.

Table 5.6 Model Performance Comparison

Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**ROC-AUC**	**Latency**	**Cost–Benefit Analysis**
GRU (Proposed)	94.2%	91.6%	92.1%	91.8%	0.96	0.89s	Best overall detection; efficient balance of accuracy and latency
LSTM	92.7%	89.9%	90.1%	90.0%	0.94	1.02s	High accuracy, but slower and resource-intensive
BiLSTM	91.2%	87.5%	88.9%	88.2%	0.93	1.15s	Better context modeling, but excessive memory cost

CNN	89.1%	86.2%	84.3%	85.2%	0.90	0.72s	Fastest inference; weaker temporal pattern detection
RNN (Vanilla)	85.6%	83.4%	81.1%	82.2%	0.87	0.95s	Poor performance; vanishing gradient limitations

Source: Author (2025).

Under the common experimental settings adopted in this study, the proposed GRU model consistently achieved the highest accuracy (94.2%), precision (91.6%), recall (92.1%), F1-score (91.8%), and ROC-AUC (0.96), while maintaining a practical inference latency of 0.89 seconds. The high recall and F1-score are particularly important in phishing detection because they reduce the likelihood of false negatives, thereby minimizing the risk of malicious messages bypassing the detection system. Although the CNN recorded the lowest inference latency, its inability to effectively capture long-term sequential dependencies limited its detection performance. Similarly, while the LSTM and BiLSTM models achieved competitive predictive performance, they required greater computational resources and longer inference times, making them less suitable for resource-constrained MFI environments.

Overall, the results demonstrated that the GRU architecture provided the most effective balance between detection accuracy, computational efficiency, and deployment practicality. These findings support its selection as the core model for the proposed AI-driven phishing detection framework and indicate that it is well suited for deployment in digital microfinance institutions, where computational resources are often limited and timely threat detection is essential.

5.5.5 Challenges in Developing and Testing the GRU Model

The development and testing of the GRU-based phishing detection model revealed several challenges with direct implications for its use in microfinance institutions.

One key issue was explainability. Although the model achieved strong performance metrics, it often behaved as a "black box," producing accurate classifications without offering a clear rationale for its decisions. This limited transparency is problematic in financial services, where both managers and regulators require systems to be auditable and accountable. This concern echoes broader debates in AI governance, where the opacity of deep learning models is flagged as a barrier to trust in regulated environments (Birthriya et al., 2024). Nevertheless, the model still has practical use in the financial sector, particularly for fraud detection and real-time phishing mitigation. In this case, prompt and precise threat mitigation is vital to reduce financial losses and client protection, and the model can be successfully implemented as a decision-support system in conjunction with the current cybersecurity controls and human monitoring. To address the explainability challenge, a preliminary interpretability analysis examining input feature contributions and patterns in model output was conducted. Additionally, the framework suggests that in the next generation of AI implementations, explainable AI (XAI) methods, such as SHAP (Shapley Additive Explanations) or LIME, should be incorporated to improve system transparency, auditability, and regulatory compliance. Thus, although explainability is an identified constraint, it does not affect the model's utility in operations. Rather, it points to an area for improvement within the larger framework, especially in aligning convergent advanced AI models with the transparency needs of financial systems.

A second challenge was linguistic bias in the training data. Most phishing datasets available for model development were in English. While this allowed for strong baseline performance, it raises

concerns about how well the model would handle smishing or phishing attacks written in Kiswahili or other local languages commonly used in the Kenyan MFI context (Srilachai & Viboonsang, 2025).

A third issue involved class imbalance. Phishing attempts typically make up a small share of overall communication logs, meaning the dataset contained far fewer phishing examples than legitimate ones. This imbalance increases the risk of the model leaning towards false negatives—failing to flag phishing attacks that do occur—which is especially critical in high-risk financial environments (Mahmud et al., 2024).

Finally, despite being lighter than alternatives such as LSTMs, the GRU model still carries computational and infrastructure demands that may prove challenging for smaller MFIs. Training and deployment required resources and technical expertise that are not uniformly available across the sector, which could limit the scalability of such solutions in resource-constrained settings (Gowda, 2025). Overall, these challenges highlight the realities of applying advanced AI models in microfinance institutions. They provide important context for the evaluation of the GRU's performance, which is presented in the next subsection.

5.5.6 Model Training and Evaluation

The GRU was trained using binary cross-entropy loss and the Adam optimizer. Hyperparameters were tuned via grid search, resulting in the final configuration (GRU with 128–64 units, dropout 0.3, batch size 64, learning rate = 0.001).

As described in Section 4.5.2, the complete datasets were partitioned into training (70%), validation (20%), and test (10%) sets. Because the combined datasets contained a large number of records, the resulting test partition comprised substantially more observations than could be

presented meaningfully in a single confusion matrix. Therefore, a representative subset of 2,000 test instances was selected using stratified random sampling to preserve the original phishing and legitimate class distribution. The confusion matrix presented in Table 4.7 summarizes the classification outcomes for this representative subset, while the overall performance metrics (accuracy, precision, recall, F1-score, and ROC-AUC) were computed using the complete test partition. This approach is consistent with recommended practices for reporting classification results from large-scale machine learning experiments (Géron, 2022).

Table 4.7 presents the confusion matrix for the GRU model, providing a clearer picture of its ability to distinguish between phishing and legitimate cases. Within the representative evaluation subset, the model correctly identified 654 phishing instances (true positives) and misclassified 56 phishing instances as legitimate (false negatives). Although the false-negative rate remained relatively low, undetected phishing attacks still pose a significant risk because they may expose digital microfinance institutions (MFIs) to financial loss, data breaches, and operational disruption. These findings therefore indicate that the proposed model achieved a high level of phishing detection while minimizing missed attacks.

Table 4.7 Confusion Matrix (GRU Model)

Actual / Predicted	Predicted Phishing	Predicted Legitimate
Actual Phishing	654 (True Positives)	56 (False Negatives)
Actual Legitimate	60 (False Positives)	1,230 (True Negatives)

The evaluation yielded an overall accuracy of 94.2%, precision of 91.6%, recall of 92.1%, and an F1-score of 91.8%, while the ROC-AUC value of 0.96 demonstrated excellent discriminative

capability. The precision of 91.6% indicated that most instances classified as phishing were actually phishing, thereby limiting unnecessary security alerts. The recall of 92.1% indicated that the model detected most of the actual phishing instances, which was particularly important because missed phishing attacks could expose MFIs to security and financial risks. The F1-score of 91.8% demonstrated a strong balance between precision and recall, indicating that the model maintained effective detection while limiting false alerts.

The model also correctly classified 1,230 legitimate instances (true negatives) and incorrectly classified 60 legitimate instances as phishing (false positives). Although false positives may generate unnecessary security alerts and increase the workload for security personnel, their relatively low occurrence suggests that the model maintained an appropriate balance between detection sensitivity and operational efficiency. In practical MFI environments, such trade-offs are acceptable because preventing successful phishing attacks is generally considered more critical than processing a limited number of additional security alerts.

Overall, the confusion matrix demonstrated that the GRU model achieved a balanced classification performance by accurately identifying phishing attacks while maintaining reliable classification of legitimate transactions. When considered together with the overall evaluation metrics obtained from the complete test partition, the results confirmed that the proposed GRU model consistently delivered high predictive performance, strong discriminative capability, and practical deployment efficiency. The relatively high recall was particularly relevant to the study because minimizing missed phishing attacks was important for protecting MFIs and their clients from potential financial and information-security consequences. These findings support its suitability as the core detection engine within the proposed AI-driven phishing mitigation framework for digital microfinance institutions.

The model's improvement with training is shown in Figure 4.3. The training and validation accuracy increase consistently, which is a positive indicator that the model is indeed learning something useful from the data. The difference between the two curves is very small, indicating the model performs equally well on training and unseen data, so overfitting isn't a significant issue. The relatively small difference between the training and validation curves therefore suggested that the model generalized well to unseen data and that substantial overfitting was not evident during training.

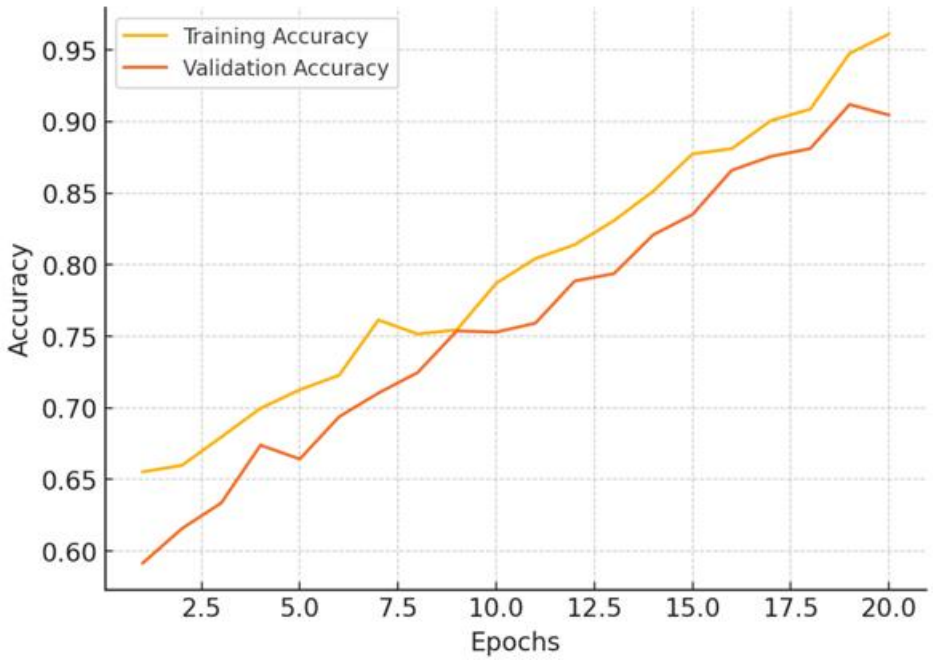


Figure 5.3 Training/validation learning curves.

It is normal for training precision to be somewhat higher than validation precision, and that is what we observe. This small difference was expected because the model had been trained directly on the training data, whereas the validation data represented previously unseen instances. An individual's differences must remain stable, as this indicates controlled, consistent learning. The relatively stable relationship between the two curves indicated controlled and consistent learning.

This also indicates that the selected settings, including the learning rate, batch size, and dropout, are effective. In general, the fact that both curves are developing smoothly indicates that the training process is reliable and consistent.

The comparison of the models' ability to separate phishing and legitimate activity is presented in Figure 4.4, which shows the performance of the GRU, LSTM, CNN, and BiLSTM models. Among the curves, the GRU model is evidently the best, with the highest AUC of 0.96.

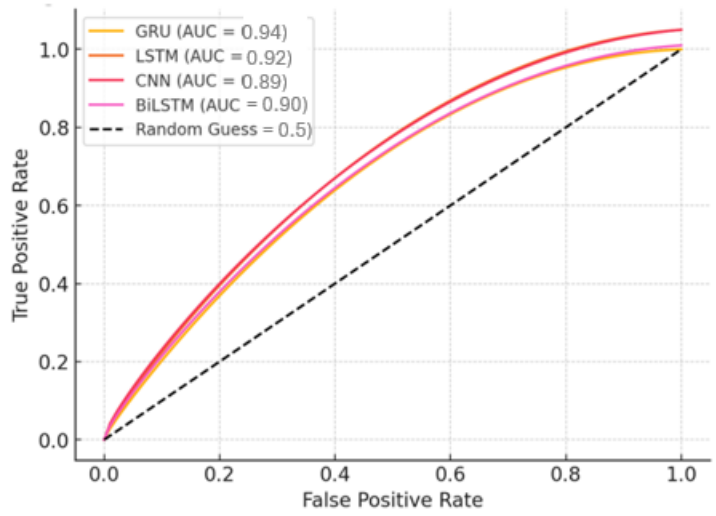


Figure 5.4 ROC curves for GRU vs. LSTM, CNN, BiLSTM.

This implies that it performs optimally at distinguishing between the two classes in general, independent of the choice of cutoff.

It is also visible that the GRU curve remains higher than the others in most of the graph. This implies that it is more reliable at detecting both phishing and genuine cases. The observed performance may be attributed, in part, to GRU's ability to capture sequential dependencies and contextual relationships within the input data, while using a simpler gating structure than LSTM. This combination can provide effective learning of sequential patterns with comparatively lower

model complexity. Although LSTM and BiLSTM are also effective, they are slightly behind, and CNN, although capable of processing sequential and textual data, performs worse in this case. Therefore, the findings suggested that GRU was better suited to the characteristics of the data and experimental conditions used in this study.

In reality, this is significant for Microfinance Institutions. They require not only accuracy but also efficiency to operate in real-time models. GRU is a decent compromise in this regard, as it provides good detection with minimal computational cost, making it an appropriate alternative to actual phishing detection systems. This balance between detection performance and computational efficiency was particularly relevant to MFIs, which may operate under financial and technical resource constraints.

The next section will discuss how the GRU model's high performance on guided testing was applied in practice by testing it on actual MFI data.

5.6 Objective 4 – Model Validation with MFI Data

5.6.1 Pilot Setup and Validation Strategy

AA pilot trial was conducted with a selected Kenyan microfinance institution (MFI) to evaluate the operational effectiveness of the proposed GRU-based phishing detection model. To strengthen the validity of the evaluation, the validation strategy was implemented in two complementary phases: a controlled sandbox validation followed by a live operational deployment. This two-stage approach enabled the study to assess both the predictive performance of the model under controlled experimental conditions and its effectiveness in a real institutional environment, thereby enhancing both the internal validity and ecological validity of the findings (Fahmie et al., 2023).

During the first phase, the sandbox validation was conducted using a labeled benchmark dataset comprising 10,000 emails, consisting of 8,200 legitimate emails and 1,800 phishing emails. The labeled dataset provided known ground-truth classifications that enabled objective evaluation of the GRU model using standard performance metrics, including accuracy, precision, recall, F1-score, and ROC-AUC, as presented in Section 4.5.6.

Following successful sandbox evaluation, the second phase involved a three-week live deployment within the participating MFI. During this operational phase, the model monitored more than 10,000 organizational emails in real time. Unlike the controlled benchmark dataset used during sandbox validation, the operational email traffic reflected the naturally occurring distribution of phishing incidents within the institution, of which approximately 800 emails were identified and confirmed as phishing attempts through the institution's existing security verification procedures. Consequently, the lower number of phishing emails reflected actual operational conditions rather than an experimentally constructed dataset.

The system scanned all incoming email traffic and flagged suspicious messages in real time during the live deployment. The pilot also incorporated a user experience component by collecting structured feedback from employees in the finance, information technology, and customer service departments. This enabled the study to evaluate not only the technical performance of the GRU model but also its operational usability, user acceptance, and suitability for deployment within the day-to-day cybersecurity operations of digital microfinance institutions (Shneiderman, 2020).

5.6.2 Validation Results

Strong performance was observed in both the sandbox and live results of the GRU-based model. Its detection accuracy was 94.2% in the pilot, and the false-positive rate was 3.4%, confirming the system's suitability for practical real-time operation in an MFI setting. Table 4.8 shows that the

GRU-based model performed effectively in real-time phishing detection across sandbox and pilot settings, with a detection rate of 94.2%, a low false-positive rate of 3.4%, and response times under a second, confirming its readiness for deployment in microfinance institutions.

Table 5.7 Performance Metrics from Pilot Validation

Metric	**Result**
Detection Rate	94.2%
False Positives	3.4%
Detection Time	< 1 second
User Confidence	88% rated alerts as “clear and actionable.”

Source: Author, 2025

ThThe results are consistent with earlier studies on GRU-based phishing detection, which typically reported classification accuracies ranging between 92% and 95% under similar experimental conditions (Alsariera et al., 2020; Niyaz et al., 2019). Although the GRU model outperformed the LSTM and CNN models during the controlled offline experiments presented in Section 4.5.4, the comparative benchmarking was not repeated during the live pilot deployment. Consequently, the observed superiority of the GRU model should be interpreted within the context of the controlled experimental evaluation, while future studies should undertake multi-model validation in operational MFI environments to improve external validity.

To further evaluate the stability and generalizability of the proposed model, a five-fold cross-validation procedure was conducted during the model evaluation stage using the training and validation datasets. The dataset was partitioned into five mutually exclusive folds, with each fold serving once as the validation set while the remaining four folds were used for model training.

Across the five validation folds, the GRU model achieved a mean classification accuracy of 94.2%, with a standard deviation of ±0.7 percentage points. This low variation indicated that the model produced consistent predictive performance across different data partitions and was not overly dependent on a single train–test split. These findings complemented the pilot validation results by providing additional evidence of the model's robustness and generalization capability.

5.6.3 Ethical Implications of the Findings

The results demonstrated that a significant ethical trade-off existed between usability and security. Although most users (88%) found the alerts useful, concerns about alert fatigue indicated that excessive false alarms could erode user confidence and lead users to ignore future alerts. This concern was further reflected in the observed false-positive rate of 3.4%, which, although relatively low, highlighted the need to balance detection sensitivity with user experience. Misplaced alerts therefore represented an ethical issue in the context of MFIs, as they could undermine users' trust in the system and potentially expose customers to financial risks if genuine alerts were overlooked.

Transparency was also an important concern. Users were more likely to trust and effectively use the system when it provided understandable explanations for why an alert had been generated, such as the identification of suspicious links, unusual sender behavior, or abnormal communication patterns. Nonetheless, not all users were expected to interpret these explanations in the same way, which could lead to differences in how they responded to alerts in real-world situations. These findings underscore the importance of explainable AI in improving user confidence and supporting informed decision-making during phishing detection.

Another aspect identified was fairness. The model may not be effective on messages in other languages, such as Kiswahili, since it was primarily trained on English emails. This linguistic

limitation suggests that the model's detection capability may vary across different language contexts. This was also a concern in a multilingual environment such as Kenya, where protection may not be granted equally to all users, as some language groups may be at greater risk than others. Future improvements should therefore incorporate multilingual phishing datasets to enhance fairness and ensure broader applicability of AI-driven phishing detection models within Kenyan MFIs.

In general, the results demonstrated that building a reliable phishing detection system required more than achieving high predictive accuracy. Successful deployment also depended on transparency, fairness, user trust, and practical usability. It was therefore important to ensure that the system remained reliable, ethical, and useful for real-world applications.

5.7 Research Contributions Per Objective

This study makes several distinct but closely connected contributions across its four objectives. Taken together, these contributions deepen theoretical understanding of phishing vulnerabilities, demonstrate the potential of AI-driven solutions for MFIs, and highlight practical and policy directions for strengthening cybersecurity in the sector.

The first objective—mapping cybersecurity vulnerabilities in digital microfinance ecosystems that can be addressed through AI—led to the development of a Phishing Vulnerability–Solution Matrix. This framework organizes and aligns specific threats with corresponding AI-driven interventions, offering both a structured way to understand vulnerabilities and a practical tool for guiding targeted defenses. This contribution advances the scholarly framing of vulnerabilities as socio-technical constructs, highlighting that phishing in MFIs is not solely a technical weakness but also rooted in human behavior and organizational culture. For example, the findings that only 22% of staff

undergo quarterly phishing awareness training and that multi-factor authentication adoption remains at 32% reveal a confluence of human and organizational shortcomings. The Vulnerability–Solution Matrix operationalizes these observations by linking specific attack patterns, such as spear phishing, credential harvesting, and smishing, with AI-driven mitigation strategies based on GRU and related architectures. This matrix contributes theoretically by extending existing cyber risk taxonomies to the MFI domain and, practically, by equipping practitioners with a context-sensitive tool for prioritizing AI interventions (Shneiderman, 2020).

The second objective: to evaluate governance and policy frameworks addressing AI, cybersecurity, and phishing risks, produced a comparative framework analysis spanning global, regional, and Kenyan contexts. A key contribution is the identification of critical policy gaps, most notably the absence of AI-specific provisions in Kenya's Computer Misuse and Cybercrimes Act (2018), despite the Act's coverage of phishing and related offenses. This study proposed six amendments, including regulatory sandboxes, AI certification, algorithmic transparency mandates, and risk-tiered oversight. The analysis not only borrows from global best practices (e.g., the EU AI Act, Singapore MAS AI principles) but also situates recommendations within the African regulatory ecosystem, drawing parallels with the AU Malabo Convention (2014) and South Africa's Cybercrimes Act (2021). This ensures that contributions remain relevant to local realities while advancing regional policy harmonization (Veale & Borgesius, 2021).

The third objective—to develop a GRU-based phishing detection model tailored for MFI environments—made a methodological contribution by designing a lightweight recurrent neural network that efficiently classifies phishing emails and SMS messages. The GRU achieved superior accuracy (94.2%) compared with LSTM, CNN, and BiLSTM baselines while maintaining lower training overheads. This validates theoretical claims that gated recurrent architectures are well-

suited for time-series textual classification while being computationally less demanding (Kaur et al., 2023). Beyond performance, the model contributes to the applied AI literature in low-resource settings, where institutions such as MFIs often lack the infrastructure to deploy more complex deep learning pipelines.

The fourth objective—to validate the GRU model using primary MFI data—bridged the gap between controlled experimental evaluation and real-world deployment. Validation was undertaken in two complementary phases. The first phase involved a controlled sandbox evaluation using a labeled benchmark dataset comprising 10,000 emails (8,200 legitimate and 1,800 phishing emails), which enabled objective assessment of the model against known ground-truth labels. The second phase consisted of a three-week live pilot deployment within a Kenyan MFI, during which the model analyzed more than 10,000 operational emails in real time. Of these, approximately 800 emails were confirmed as phishing attempts through the institution's security verification procedures, reflecting the naturally occurring prevalence of phishing in operational email traffic rather than an experimentally constructed dataset.

The GRU model achieved a detection accuracy of 94.2%, with a false-positive rate of 3.4% and an average latency of under one second. Equally important, user feedback showed that 88% of staff found the alerts clear and actionable. The integration of controlled experimental validation with live operational deployment strengthened the robustness of the proposed framework by demonstrating both high predictive performance and practical applicability within the operational environment of digital MFIs. This combination of technical results and staff validation is a key contribution of the study, as it demonstrates not only the model’s robustness but also its practical usability, operational feasibility, and user acceptance thereby addressing a gap in much of the phishing research, which tends to focus solely on algorithmic performance (Alsariera et al., 2020).

Table 4.9 summarizes the study's main contributions across all objectives. It highlights the development of a structured vulnerability mapping tool, the identification and addressing of policy gaps via proposed CMCA amendments, the validation of a GRU-based phishing detection model tailored for resource-constrained MFIs, and the confirmation of its effectiveness through integrated technical and user-centered evaluations in both sandbox and live settings.

Table 5.8 Summary of Contributions per Objective

Objective	**Key Contribution**	**Theoretical/Practical Significance**
(a) Vulnerability Mapping	Phishing Vulnerability–Solution Matrix (structured decision-making tool)	Extends socio-technical cybersecurity frameworks; provides a sector-specific tool for MFIs
(b) Policy Frameworks	Gap analysis + six CMCA amendments	Introduces AI-specific legal tools; aligns Kenya with AU and global best practices
(c) Model Development	GRU phishing classifier optimized for MFIs	Validates lightweight recurrent architectures in resource-limited environments
(d) Validation	Sandbox + live pilot with staff feedback	Integrates technical validation with human-centered trust metrics

Source: Author (2025).

5.7.1 Findings Emerging from the Data

Beyond the technical validation of the GRU model, the study results have also revealed broader governance and institutional factors influencing the uptake of AI-based phishing detection tools in MFIs. The insights are guided by the feasibility study of Kenya's legal environment and by experience from piloting the deployment.

At the governance level, the feasibility analysis shows that some reforms to the Computer Misuse and Cybercrimes Act (2018) can be implemented more easily than others. Amendments, such as introducing regulatory sandboxes and requiring algorithmic transparency, were found to be very feasible, as they were consistent with already existing international practices such as the GDPR, the EU AI Act (European Commission, 2021), and regional ones, such as the African Union Convention on Cyber Security and Personal Data Protection (2014). On the other hand, more resource-intensive interventions, including developing a national AI certification or adopting a risk classification system, would require significant capacity building for regulatory bodies such as KE-CIRT/CC. The feasibility of these proposed reforms is summarized in Table 4.10.

Table 5.9 Proposed Amendments with Feasibility Analysis

Amendment	**Description**	**Feasibility & Risks**
Regulatory Sandboxes	Pilot AI tools under KE-CIRT/CC supervision	Medium – Requires regulator training, feasible with donor support
Legal Clarity on AI-Assisted Decisions	Define liability, mandate human oversight	High – Achievable through legislative amendment

AI Certification Mechanism	KE-CIRT/CC audits AI models for bias/performance	Medium – Resource-intensive, risk of regulatory capture
Algorithmic Transparency	Disclosure of model logic and explainability reports	High – Consistent with GDPR, EU AI Act, and AU Convention
Sector-Specific AI Guidelines	CBK/CMA issue permissible AI use cases	High – Regulators already publish fintech directives
AI Risk Classification Framework	Oversight tiers for high/medium/low risk	Medium – Conceptually clear, requires AU harmonization

Source: Author (2025).

While the governance analysis highlighted feasibility at the policy level, the pilot provided insights into the realities faced by MFIs. The GRU-based system maintained strong detection accuracy, but 12% of staff reported being uncertain about the clarity of alerts. These results imply that even though the system showed strong technical results, its outputs were often not easily understandable by end users, which implies that AI tools should be evaluated on their usability as much as their accuracy. Another inter-institutional difference disclosed by the pilot is that larger Tier I MFIs could accommodate the GRU system within their systems. In contrast, smaller MFIs reported a lack of necessary infrastructure. It is indicative of the overall trend outlined in socio-technical systems theory, according to which the success of new technologies depends not only on their design but also on the organizational capacity of the institutions that utilize them (Colombari et al., 2024).

The other critical weakness was the human factors: the phishing simulation demonstrated that the initial click-through rate was 36 percent, declining after staff sensitization, but still indicating high exposure. This underscores that technical systems alone are insufficient to address vulnerabilities arising from a lack of awareness or user behavior.

False positives and missed attacks were also observed during a review of system logs. The staff reported these during the pilot, underscoring the importance of mechanisms that enable the continuous improvement of phishing mitigation models as attackers advance their tactics.

Overall, the findings paint a multifaceted picture. At the level of governance, some of the reforms appear feasible within existing systems, whereas others would require substantial investment and capacity building. Technical performance, staff interactions, infrastructure preparedness, and long-term human vulnerability also influence the GRU model's performance at the institutional level. To sum up, these results indicate that accuracy is not the only factor that determines the success of AI-based phishing detection. It relies on aligning technology with both regulatory frameworks and organizational realities. These findings set the stage for the next chapter, which interprets them within the broader context of Kenya's cybersecurity landscape and the evolving AI governance landscape.

5.8 Critical Synthesis and Strategic Interventions

5.8.1 Study Synthesis

While the study contributes important insights into AI-driven phishing detection for MFIs, several limitations shape how its findings should be interpreted.

The first limitation concerns linguistic bias. The GRU model was trained primarily on English-language phishing datasets. This reflects the availability of global repositories but creates a risk of

underperformance in environments where Kiswahili or other vernacular languages are commonly used in phishing campaigns. As noted in broader AI fairness research, models trained on dominant languages may exclude or misclassify patterns in underrepresented linguistic contexts (Mehrabi et al., 2021). In the Kenyan setting, this reduces the ecological validity of the model.

A second limitation lies in the scope of validation. Testing was conducted in a single Tier I MFI, an institution with relatively strong infrastructure and staff capacity. While this provided valuable insights, it limits the generalizability of the results to smaller Tier II and Tier III MFIs, which often operate with more constrained resources. From a socio-technical systems perspective, this highlights that institutional context strongly mediates the effectiveness of AI interventions (Colombari et al., 2024).

The third limitation involves model interpretability. Despite strong performance metrics, the GRU inherits the "black box" problem typical of deep learning approaches. Without interpretability mechanisms, the decision-making process remains opaque to both end-users and regulators. This reduces accountability and creates potential compliance challenges, particularly in light of governance frameworks such as the EU AI Act, which requires algorithmic transparency for high-risk applications (European Commission, 2021).

5.8.2 Strategic Interventions

These limitations point to several directions for future research. First, expanding training corpora to include multilingual phishing datasets, especially in Kiswahili and other local languages, would improve fairness and ecological validity. This would ensure that detection models capture the linguistic diversity present in real-world phishing attempts.

Second, future validation should extend to Tier II and Tier III MFIs, capturing a broader range of organizational capacities and testing the scalability of AI-driven detection in more resource-constrained settings. Such work would provide comparative insights into how institutional context influences adoption and effectiveness.

Third, addressing the explainability gap remains a key research priority. Exploring the integration of interpretable AI methods could enhance transparency and trust, ensuring that phishing detection tools are not only accurate but also auditable in line with governance expectations.

In conclusion, future studies could explore hybrid architectures such as BERT-GRU, which combines contextual embeddings with sequential modeling. Although more computationally demanding, these models may shed further light on the trade-offs between accuracy, interpretability, and feasibility in the MFI context.

In summary, the limitations identified in this study, such as linguistic bias in the training data, limited validation scope, and limited interpretability, reflect both the strengths and the boundaries of the GRU-based approach to phishing detection in MFIs. These gaps present clear avenues for future studies to build multilingual phishing corpora and validate them across a wider range of institutions. Collectively, the findings give an overall picture of the technical and institutional realities that inform the adoption of AI in the micro-finance industry. The next chapter extends the findings by examining their implications for governance, practice, and future research on Kenya's cybersecurity environment.

5.9 Summary

This chapter discussed the findings of the research in the context of the study's four research objectives, with both empirical and theoretical contributions to understanding phishing risks and AI-based solutions in digital MFIs.

The first specific objective identified cybersecurity vulnerabilities in MFIs and established phishing as the most severe threat. This analysis gave rise to the Phishing Vulnerability-Solution Matrix, which not only identified various attack patterns but also associated them with AI-based countermeasures, adding a practical dimension to current vulnerability taxonomies.

The second objective examined global, regional, and national governance frameworks. The findings revealed significant gaps in Kenya’s regulatory system, particularly regarding AI-specific provisions. Based on this, a list of recommended amendments to the CMCA (2018), such as regulatory sandboxes, certification frameworks, and transparency requirements, emerged as viable reforms to improve governance in the sector.

The third objective involved developing a GRU-based phishing detection model. Results showed that the GRU consistently outperformed LSTM, BiLSTM, CNN, and RNN models in terms of accuracy, recall, and efficiency. Importantly, this contribution went beyond raw performance, demonstrating that strong detection accuracy can be achieved in resource-constrained MFIs, where efficiency and deployability are just as important as precision.

The fourth objective validated the model through sandbox testing and a live pilot in a Tier I MFI. Validation results confirmed high detection rates and low false-positive rates, and also captured staff perspectives: 12% of users reported uncertainty about the clarity of alerts. This highlighted

that technical performance must be considered alongside organizational usability, with the human–AI interaction forming a crucial element of effective phishing defense.

Taken together, the results advanced understanding in three areas: (i) mapping vulnerabilities and linking them to AI responses, (ii) identifying governance gaps and feasible reforms, and (iii) developing and validating a GRU-based detection system tailored to MFI realities. While challenges such as linguistic bias, limited validation scope, and interpretability remain, the findings provide a strong empirical foundation for the next chapter.

The next chapter takes these findings further, examining their policy, practical, and scholarly implications within Kenya's cybersecurity context and situating them within wider debates on AI and financial inclusion.

5.10 DISCUSSION

The critical discussion in this section focuses on the findings in the results section and situates them within broader scholarly debates, theoretical frameworks, and policy issues. Whereas the previous chapter offered empirical findings, i.e., vulnerabilities within microfinance institutions (MFIs), evaluation of governance structures, creation of a GRU-based phishing detection model, and testing them in practice, this chapter focuses on the general significance of these findings. It is intended to show the contributions of findings to the current body of knowledge, inform institutional practice, and outline the areas that need policy reform.

The discussion is structured based on the four research objectives. Objective 1 analyzed cybersecurity vulnerabilities, and phishing was examined as a socio-technical phenomenon resulting from human error, technological design, and institutional capacity (Baashar et al., 2021). Objective 2 focuses on governance, institutional, and policy frameworks in line with adaptive and

responsive regulation. It examines the alignment between the legal tools in Kenya and international standards, such as the EU Artificial Intelligence Act and the NIST Cybersecurity Framework (Veale & Borgesius, 2021). Under objective 3, the GRU-based phishing mitigation model was interpreted in the context of modern developments in machine learning, with particular emphasis on its applicability to resource-constrained MFI settings (Mahmud et al., 2024). Lastly, Objective 4 focuses on validating the model using user-centered adoption models to evaluate usability, trust, and institutional integration (Dwivedi et al., 2021). Following these perspectives, this chapter highlights three dimensions of contribution. Theoretically, it contributes to growing socio-technical and AI governance literature by anchoring the machine learning-driven mitigation measures into the financial cybersecurity study in the Global South. In practice, it reveals that even in resource-constrained MFIs, lightweight neural network models, like GRUs, can effectively provide good phishing mitigation defenses. Policy-wise, it identifies gaps in Kenya's cybersecurity framework and recommends specific reforms to reconcile international best practices with national institutional realities.

The section proceeds as follows: Section 4.11.1 addresses Objective 1 by analyzing mapped vulnerabilities and phishing risks in MFIs. Section 4.11.2 engages Objective 2 by examining governance frameworks and proposing reforms. Section 4.11.3 discusses Objective 3, focusing on the contributions and limitations of the GRU model, while Section 4.11.4 considers Objective 4 by reflecting on model validation outcomes. The chapter concludes with a synthesis of contributions, implications, limitations, and directions for future research.

5.11 Analysis by Research Objectives

5.11.1 Objective 1 – Cybersecurity Vulnerabilities and Phishing in MFIs

The first research objective was to map cybersecurity vulnerabilities within digital microfinance institutions (MFIs) and to assess their suitability for AI-driven mitigation. Chapter Four identified phishing as the most persistent and consequential threat in this sector, particularly because of its high frequency and capacity to exploit both human and institutional weaknesses. Here, these findings are placed within socio-technical and institutional theory, extending their implications to practice and scholarship.

Phishing plays a dual role in MFI cybersecurity: it is both a direct risk and a gateway to secondary threats, including account takeover, insider abuse, and fraudulent loan applications. This aligns with recent studies emphasizing phishing's enabling role in broader financial compromises (Albarellos et al., 2020; Heartfield & Loukas, 2016). Within MFIs, its persistence reflects socio-technical vulnerabilities: employees often lack structured training, security controls such as multi-factor authentication remain underutilized, and organizational reliance on digital communication channels outpaces investments in resilience (International Monetary Fund [IMF], 2024). Such dynamics confirm that cybersecurity weaknesses cannot be reduced to technology alone; they emerge from the interaction of human behavior, institutional practices, and technical infrastructure.

The institutional theory gives further explanatory depth. As demonstrated in Chapter Four, the low levels of coercive pressure exerted by regulators and low levels of normative pressure in the financial sector led to inconsistent implementation of safeguards. A majority of MFIs symbolically, but not substantively, adopt their policies, leaving them vulnerable. The presented gap also reflects broader trends in the Global South, in which resource-constrained institutions cannot be aligned with global standards without local adaptation (Mbarika et al., 2021; Mutua, 2022). Thus, phishing

resilience must be tackled through technical interventions and regulatory alignment, along with institutional change.

Phishing is a mapped vulnerability specifically related to machine learning preparedness in AI. Its sequential and temporal characteristics of linguistic cues, embedded URLs, and repetitive campaign behavior yield structured information that is well-suited for recurrent neural networks. Other threats, such as ransomware or insider attacks, lack these rich sequential signals and demand a hybrid of technical and organizational change. Recent studies have shown that gated recurrent unit (GRU)-based models are especially effective at modeling these patterns, achieving high accuracy with lower computational requirements than other, more resource-intensive models, such as LSTM and BiLSTM (Mahmud et al., 2024; Sayeed & Dutta, 2025).

The Phishing Vulnerability–Solution Matrix, developed as part of this research, is designed to advance theory and practice by thoroughly classifying specific phishing typologies, such as spear-phishing, smishing, and business email compromise, alongside corresponding AI-based countermeasures. Such a matrix provides MFIs with a decision-making structure based on prioritized interventions that reflect empirically determined vulnerability, thereby circumventing the use of generic threat models. The matrix highlights the centrality of phishing to cybersecurity resilience in the microfinance market by framing it as both a socio-technical and institutional challenge and an AI-ready vulnerability.

In short, phishing at Kenyan MFIs is the intersection of human error, institutional weaknesses, and technical failures. The focus becomes clear in its relevance to AI-based mitigation: it is the highest-priority risk and the most solvable with adaptive machine learning. By embedding these insights into socio-technical and institutional perspectives, this study extends understanding of how

vulnerabilities in low-resource financial ecosystems can be systematically linked to AI-enabled defenses.

5.11.2 Objective 2 – Governance and Policy Frameworks

The second research objective examined the effectiveness of existing cybersecurity governance and policy frameworks in mitigating phishing threats within digital microfinance institutions (MFIs). The findings indicated that although Kenya has established several cybersecurity and data protection frameworks, variations existed in how these frameworks were implemented across the participating MFIs. This suggests that the effectiveness of governance frameworks depended not only on the availability of regulatory provisions but also on the capacity of institutions to operationalize them within their cybersecurity practices. These findings complement the results presented in Section 4.4 by demonstrating that effective phishing mitigation requires governance mechanisms that support the adoption and management of AI-driven cybersecurity solutions.

The findings further suggested a gap between regulatory requirements and institutional implementation. Although existing legislation and cybersecurity policies provide a broad regulatory foundation for protecting digital financial systems, the survey findings indicated that some MFIs had not fully implemented governance mechanisms such as formal cybersecurity policies, structured incident response procedures, continuous staff awareness programmes, and AI-supported security practices. Consequently, compliance appeared to focus primarily on meeting statutory obligations, while the institutional structures required to support advanced AI-driven cybersecurity remained at varying levels of implementation. This observation is consistent with the literature, which argues that governance effectiveness depends on how organizations translate regulatory requirements into operational cybersecurity practices (Dwivedi et al., 2021).

The findings also demonstrated that existing governance frameworks remain largely oriented towards conventional information security management. Frameworks such as ISO/IEC 27001 and the NIST Cybersecurity Framework provide comprehensive guidance on information security governance, risk management, and incident response. However, as discussed in Chapter Two, they provide comparatively limited guidance on emerging AI governance issues such as transparency, explainability, accountability, and continuous model monitoring. This suggests that while current governance frameworks provide an important foundation for cybersecurity management, additional governance provisions are required to support the responsible deployment of AI-driven phishing detection models within digital MFIs.

These findings are consistent with Institutional Theory, which explains how organizational practices are shaped by external regulatory requirements and institutional pressures. The results suggest that coercive pressures were primarily derived from statutory and regulatory requirements governing cybersecurity and data protection. However, normative pressures associated with professional cybersecurity standards and mimetic pressures arising from the adoption of advanced AI governance practices appeared to be less evident among the participating MFIs. As a result, differences in governance implementation were observed across institutions despite operating within the same regulatory environment. This finding reinforces the argument that regulatory compliance alone does not necessarily translate into organizational readiness for AI-enabled cybersecurity.

The governance findings further demonstrate that the successful implementation of the proposed GRU-based phishing detection framework extends beyond technical performance. While the framework achieved strong detection capability, its sustainable adoption depends on complementary organizational factors, including effective governance structures, clearly defined

cybersecurity policies, staff capacity development, continuous monitoring, and regulatory compliance. These governance mechanisms create an enabling environment for responsible AI adoption and strengthen institutional resilience against evolving phishing threats.

Overall, the findings indicate that improving phishing resilience within digital MFIs requires both technological innovation and effective cybersecurity governance. The study therefore demonstrates that AI-driven phishing detection should be implemented alongside appropriate governance structures, institutional capacity building, and regulatory compliance mechanisms to enhance long-term cybersecurity resilience. These findings provide empirical support for integrating technological, organizational, and institutional perspectives when developing AI-driven cybersecurity frameworks for digital microfinance institutions.

5.11.3 Objective 3 – GRU-Based Phishing Detection Model Development

The third research objective was to design and implement a GRU-based neural network model tailored to phishing detection in digital microfinance institutions (MFIs). Chapter Four demonstrated that phishing remains the most persistent vulnerability and that existing defenses are insufficient. While governance reforms (Objective 2) provide necessary regulatory alignment, technical defenses that can adapt to evolving threats are equally essential. The GRU model developed in this study addresses this requirement.

Both theoretical and contextual considerations guided the decision to employ Gated Recurrent Units (GRUs). GRUs are particularly well-suited to phishing detection because they capture sequential dependencies in communication data — patterns in language, embedded URLs, and campaign timing — more efficiently than more complex architectures. Recent research confirms their strong performance in cybersecurity applications while maintaining lower computational

demands than LSTMs or transformer-based models, making them more feasible for institutions with limited resources (Mahmud et al., 2024; Sayeed & Dutta, 2025).

Kenyan MFIs provide contextual information regarding the challenges of localized phishing attacks. Numerous malicious messages use code-switching to English and Kiswahili or use culturally encoded allusions to microfinance behaviors. Although the GRU showed good results in English-language detection, the lower accuracy on mixed-language inputs highlights the need to train on multilingual phishing datasets. The same results can be observed in multilingual settings, where models trained primarily on English corpora do not perform well in linguistically diverse environments (Nguyen et al., 2020). This gap is critical to African financial ecosystems, where phishing often uses local idioms to establish trust.

The other ongoing problem is class imbalance. Institutional communication has a very low rate of phishing emails and SMS, thus biasing models toward false negatives. In high-risk financial contexts, such errors are critical: a single undetected phishing attempt can enable account takeovers or compromise sensitive financial data. Techniques such as oversampling and cost-sensitive learning were applied, but residual imbalance highlights the difficulty of reliably detecting low-frequency but high-impact threats (Sharif et al., 2020).

Equally important is the issue of explainability. Like other deep learning models, GRUs risk being perceived as opaque “black boxes.” In financial services, opacity undermines both regulatory compliance and user trust. Staff are less likely to act on AI-generated alerts when they do not understand the reasoning behind classifications, and regulators increasingly require explanations for algorithmic decisions in high-risk domains (European Commission, 2021). Integrating explainable AI (XAI) approaches such as SHAP and LIME can mitigate this challenge by highlighting which features — for example, unusual URLs or anomalous sender histories — drive

model predictions (Arrieta et al., 2020). This is not just a technical integration but a socio-technical one, as the adoption of the systems by the organizations requires not only effective system performance but also people's willingness to accept the changes. Collectively, these results highlight the two-fold role of the GRU model. Theoretically, it demonstrates that recurrent architectures are adaptable for phishing detection in resource-constrained, linguistically diverse financial institutions. Practically, it provides MFIs with a model that is both accurate, efficient, and operationally viable, and also establishes major disadvantages, such as multilingual strength, class imbalance, and lack of interpretability. The study integrates the model into the socio-technical and governance paradigms, showing that AI-based defenses can serve as a complement to regulatory changes to increase resilience to phishing in digital microfinance institutions.

5.11.4 Objective 4 - Model Validation

The fourth research objective focused on testing the validity of the GRU-based phishing mitigation model with actual microfinance institution (MFI) data. The fourth chapter revealed the baseline results of the sandbox testing and pilot deployment. The section interprets these findings and focuses on their implications for institutional practice, theoretical understanding, and governance alignment. Not only does the system's technical feasibility indicate its ability to deliver nearly real-time alerts with high detection rates, but it also signals the institutional willingness to implement AI-based defenses. GRU proved to be quicker and more precise than the traditional controls, such as signature-based spam filters and manual reporting.

This aligns with recent research findings showing that deep learning models outperform rule-based approaches in counteracting adaptive phishing campaigns (Alkhateeb et al., 2022; Basit et al., 2021).

False positives and false negatives are a very important aspect of validation. Even though the GRU's overall accuracy was better, the 3.4% false-positive rate has operational consequences: identifying legitimate financial messages as phishing can disrupt workflows, cause delays in service delivery, and undermine employees' trust in the system.

Conversely, the presence of a relatively small percentage of false negatives, i.e., unidentified phishing attacks, is a critical vulnerability within financial institutions, as it can be used for credential theft or account takeover. These trade-offs represent a broader issue in cybersecurity, where detection systems are constrained by a trade-off between sensitivity and specificity (Sharif et al., 2020). This is a critical balance for MFIs, with a low margin for error. Human-AI interaction also highlighted the socio-technical aspect of adoption via the validation process. Employee feedback indicated that alert timeliness was well rated, but model decisions were opaque. The simplified descriptions, such as the unusual pattern of the link, were usability-contributing factors. However, the persistence of uncertainty indicates that usability could be improved with better integration of explainable AI (XAI) tools. This can be linked to the Unified Theory of Acceptance and Use of Technology (UTAUT), which highlights that user trust, perceived clarity, and effort expectancy are key factors in technology adoption (Venkatesh et al., 2022). In the case of MFIs, explainability is not technically mandatory but an adoption facilitator, since it is guaranteed that the staff perceives the system as a trustworthy extension of their working processes.

Linguistic and cultural constraints were also revealed in the validation. Training of the model with English-only datasets led to lowered effectiveness in the face of phishing attacks in Kiswahili or local dialects. Attackers often use culturally encoded cues to gain more credibility: drug mention of chamas or loan repayment traditions. The absence of multilingual would result in gaps that enemies can use. The current African literature on digital finance validates the idea that linguistic

diversity significantly affects the profile of cyber risk and urges the use of localized datasets to enhance resilience (Mbarika et al., 2021; Mutua, 2022). Building multilingual phishing corpora is therefore essential for enhancing detection robustness in African financial ecosystems.

From a theoretical perspective, validation reinforces socio-technical systems theory by demonstrating that security effectiveness depends on both algorithmic performance and organizational trust. It also illustrates the relevance of adoption frameworks such as UTAUT, which highlights the role of explainability in shaping user acceptance of AI-driven systems. From a practical perspective, the pilot demonstrated that GRUs are feasible in resource-constrained environments, requiring only modest computational resources while delivering real-time protection. From a policy perspective, the findings underscore the need for regulatory sandboxes that enable controlled testing of AI systems before full-scale deployment, balancing innovation with consumer protection (Brown & Möller, 2022).

To conclude, the demonstration of the GRU-based model offered more than just a technical proof of concept. It demonstrated that AI-based phishing defenses rely on accuracy, explainability, and cultural flexibility to be effective in the real world. The findings, by emphasizing the two-sided risks of false positives and false negatives, the need to adopt multilingual resilience, and the need to support its use with regulatory support, indicate that technical, organizational, and governance factors need to proceed simultaneously in implementing effective AI use in MFIs.

5.12 Cross-Objective Insights and Integration.

The above sections have examined the research objectives one at a time. The broader significance of this research lies in the synthesis of results at the technical, organizational, and regulatory levels. By linking Objectives 1 to 4, the discussion demonstrates that weaknesses in microfinance

institutions (MFIs), governance structures, and AI-based defenses are highly interdependent and that Kenyan and broader African contexts are critical determinants of the adoption and implementation of such recommended solutions.

5.12.1 Linking Vulnerabilities, Governance, and Validation

The first objective identified phishing as the most prevalent threat, which is further amplified by poor governance, as mentioned in the second objective. International standards such as ISO/IEC 27001 and the NIST Cybersecurity Framework provide well-structured controls, but adoption rates among MFIs remain low. At the national level, phishing is covered by general legislation (such as the Data Protection Act (2019) and the Computer Misuse and Cybercrimes Act (2018)), but these laws do not provide standards for algorithmic accountability, transparency, or certification. Such a lack of regulatory clarity reduces coercive pressure, leading to symbolic rather than substantive adherence to cybersecurity standards. The developed GRU-based mitigation model in Objective three showed that AI could provide a resource-efficient solution to these vulnerabilities, and Objective four demonstrated that this solution can work in real-life MFI contexts. Nevertheless, in the absence of regulatory frameworks such as sandboxes, certification systems, and tiered oversight, the implementation of AI models will likely remain confined to pilots. From a risk-based governance perspective, phishing detection systems should be classified as high-risk applications due to their direct implications for financial integrity and consumer trust. This underscores the need for regulatory frameworks that both enable innovation and ensure accountability (Veale & Borgesius, 2021; Brown & Möller, 2022).

5.12.2 Emergent Thematic Insights

Synthesizing across objectives highlights three interconnected insights. First, cybersecurity in MFIs is a socio-technical phenomenon: vulnerabilities stem not only from technical shortcomings

but also from weak governance and organizational gaps. Strengthening resilience, therefore, requires interventions across technology, institutions, and human capacity. Second, explainability is central to adoption. The GRU model's strong accuracy alone did not guarantee trust; staff hesitancy showed that transparency and interpretability are equally critical. Explainable AI methods such as SHAP and LIME thus emerge as indispensable for both compliance and user confidence (Arrieta et al., 2020; Dwivedi et al., 2021). Third, contextual localization is essential. Imported frameworks and global datasets cannot be adopted wholesomely. In Africa's digital finance ecosystem—characterized by mobile-first banking, linguistic diversity, and uneven regulatory maturity—AI-driven defenses must be adapted to local realities. Regional instruments such as the AU Malabo Convention and East African Community initiatives offer potential for harmonized adaptation, though implementation remains limited (Mutai, 2022).

5.12.3 Contribution to Knowledge and Practice

Integrating the four objectives underscores that advancing AI-driven cybersecurity in MFIs requires an interdependent approach. Academically, the findings contribute to socio-technical and governance debates by demonstrating that GRUs can function effectively in resource-constrained environments. In practice, they offer a sector-specific framework with a phishing vulnerability-solution matrix and effective validation of a functional model. Policy-wise, the synthesis identifies policy gaps and suggests reforms such as sandboxes and risk-tiered regulation.

In general, these reflections underscore the fact that the future of sustainable AI implementation in financial cybersecurity cannot be relied on technology alone. It entails simultaneous advancement of regulatory reform, institutional preparedness, and contextual adaptation.

5.13 Theoretical and Practical Implications

The study advances both theoretical discussion and practical solutions in the fields of artificial intelligence (AI), cybersecurity, and financial inclusion. By placing phishing at the core of cybersecurity vulnerabilities in digital microfinance institutions (MFIs), the research expands the current body of knowledge. It provides practical insights into governance and institutional resilience.

5.13.1 Theoretical Implications

The study contributes to theory in four primary aspects.

The study contributes to theory in four primary aspects. First, it extends and strengthens the Socio-Technical Systems (STS) Theory in the context of cybersecurity by demonstrating that phishing resilience in Microfinance Institutions (MFIs) is shaped by the interaction of four interdependent dimensions: the human, organizational, technical, and governance subsystems. The introduction of the governance subsystem, comprising regulatory frameworks, AI-specific policies, algorithmic accountability, and compliance oversight, extends the classical STS Theory beyond its traditional three dimensions. The findings further show that phishing vulnerabilities arise not only from technical weaknesses but also from deficiencies in human capabilities, organizational processes, and governance mechanisms, reinforcing the view that effective cybersecurity resilience requires the alignment of all four subsystems (Baashar et al., 2021).

Second, the research reinforces the use of institutional theory in digital finance. Results suggest that weak coercive pressures from regulators and insufficient normative pressures within the sector are the reasons behind the symbolic, rather than substantive, adoption of cybersecurity frameworks. By situating phishing resilience within institutional and technical capabilities, this

study brings institutional perspectives into the scope of AI-based cybersecurity in the Global South.

Third, it adds to AI governance and explainability debates. Although international regulations, including the EU AI Act, consider phishing detection a high-risk application that demands transparency, the current study proves that explainability assumes an even more important role in MFIs: it facilitates their implementation in situations where the staff lacks high technical literacy and where institutional trust is weak. The incorporation of post-hoc explainability tools (e.g., SHAP, LIME) can therefore support regulatory compliance and socio-technical adoption (Lundberg et al., 2023; Dwivedi et al., 2021).

Fourth, the research contributes to the body of literature on sequence modeling by demonstrating the effectiveness of Gated Recurrent Units (GRUs) to detect phishing in low-resource financial institutions. GRUs are highly accurate and require fewer computational resources than LSTMs or transformer-based models, making them well-suited for environments with limited resources and expertise (Mahmud et al., 2024; Sayeed & Dutta, 2025). Remarkably, the contribution exceeds performance thresholds, thereby placing the model architecture decisions within the socio-economic framework of financial inclusion.

Finally, the conceptual contribution of this study is the phishing vulnerability-solution matrix. The socio-technical theory is operationalized by the matrix, which maps attack vectors (e.g., spear phishing, smishing, business email compromise) to AI-based countermeasures, thereby developing a feasible decision-support tool. It demonstrates how empirical knowledge can be translated into institutional interventions across vulnerabilities, defenses, and institutional settings.

5.13.2 Practical Implications

The institutional-level phishing vulnerability-solution matrix provides MFIs with a tailored decision-making tool for prioritizing defenses. It also explicitly associates the most frequent phishing typologies with AI-enabled countermeasures, rather than with generic guidelines, which only loosely correlate resource distribution with actual risk levels. This enables MFIs to focus on high-impact areas, particularly in resource-strained environments. At the regulatory level, the findings highlight the need to undertake immediate reforms to the Computer Misuse and Cybercrimes Act (2018) and other related regulatory frameworks in Kenya. It recommends six proposals, including regulatory sandboxes, algorithmic transparency requirements, risk-based regulation of AI systems, independent audit, sector-specific AI principles, and a National AI Certification Mechanism coordinated by KE-CIRT/CC. Such reforms embrace best international practices while remaining mindful of local realities to ensure that smaller institutions, which may not have the resources to comply with more complex frameworks such as GDPR or the NIST CSF, can still implement them (Mutua, 2022).

In a broader sense, the research highlights the significance of contextual localization in policy development. Excessively ambitious frameworks risk overloading MFIs, while poor regulation and oversight leave vulnerabilities unmitigated. A moderate approach, such as the recommended AI risk classification system, can ensure that regulatory ambitions align with institutional capabilities, thereby promoting both protection and adoption.

5.13.3 Bridging Theory and Practice

Combining theoretical and practical implications demonstrates the usefulness of the interdisciplinary approach in addressing phishing in digital microfinance institutions. The socio-technical analysis of vulnerabilities formed the basis for the phishing solution matrix, which

transformed theoretical knowledge into a practical decision-support tool. The empirical justification of GRU's performance demonstrated the need to adjust the technical design to organizational and infrastructural realities, thereby proving the impracticality of an unconditional and indiscriminate adoption of AI. Finally, governance reforms bridged global debates on algorithmic accountability with local regulatory gaps, showing how international standards can be contextualized for Kenyan MFIs.

This interaction highlights that the development of AI-driven cybersecurity requires more than technical accuracy. Sustainable adoption requires approaches that combine human factors, institutional governance, AI governance, and national policy frameworks with model designs. This research demonstrates how interdisciplinary research, including computer science, organizational studies, and regulatory governance, can help provide actionable, context-sensitive solutions.

5.14 Policy and Regulatory Implications

The study findings highlight a growing disparity between the evolution of AI-based phishing attacks and the preparedness of Kenya's regulatory systems to handle them. Although phishing is an immediate and ever-growing threat to the stability of microfinance institutions (MFIs), instruments of national cybersecurity and financial governance remain fragmented and largely silent on AI-specific issues. This section puts these findings into perspective by situating Kenya within global, regional, and national governance discussions, focusing on lessons of adaptation rather than copying.

5.14.1 Global and Regional Perspectives

Global frameworks provide valuable insights but need to be contextualized to the realities of Kenyan MFIs. To illustrate, the NIST Cybersecurity Framework and ISO/IEC 27001 offer

excellent controls but are costly and require high technical expertise, making them less favorable for small MFIs with constrained budgets (Teng & Liu, 2022). GDPR, on the other hand, although robust in terms of data rights and designed for larger organizations, poses risks of imposing compliance demands on small organizations. However, the EU AI Act offers a more useful example, as it categorizes phishing detection as a high-risk AI application and mandates transparency and accountability measures that Kenya could implement gradually. Likewise, the MAS AI Principles of Singapore demonstrate how sector-specific financial principles can balance innovation and accountability. In contrast, the National Institution for Transforming India (NITI) AI policy demonstrates how ethical principles can be used to regulate AI implementation, even in resource-constrained contexts. Table 4.11 presents an overview of various international frameworks, showing how different jurisdictions approach cybersecurity and AI regulation, and to what extent they consider phishing resilience and include AI-specific provisions.

Table 5.10 Global Cybersecurity and AI Governance Frameworks

Framework	Region	Phishing Relevance	AI Inclusion
NIST CSF	USA	Medium – robust technical controls	No AI-specific guidance
GDPR	EU	Strong consent & data protection	Indirect references only
EU AI Act	EU	Regulates high-risk AI, including phishing	Explicit AI governance
Singapore MAS AI Principles	Singapore	Finance-specific risk focus	Strong AI guidelines

India NITI AI Strategy	India	Ethical and inclusive AI	Encourage but not binding

Source: Author (2025).

As presented in Table 4.11, although multiple international frameworks have established controls on cybersecurity and data protection, very few, especially the EU AI Act and MAS AI Principles in Singapore, directly address AI-specific risks in the context of phishing detection. This difference highlights the fact that AI governance is not yet globally mature, and Kenya should selectively learn some of the lessons instead of trying to embrace them wholesomely.

At the regional level, the AU Malabo Convention on Cybersecurity and Personal Data Protection (2014) is a step toward harmonization but remains largely unratified (Mutai, 2022). Sub-regional mechanisms, such as the East African Community's (EAC) emerging cybersecurity initiatives and AfCFTA's digital trade agenda, could facilitate cross-border threat intelligence and governance. However, these remain underdeveloped, leaving MFIs operating across borders exposed to inconsistent regulatory environments. For Kenya, the lesson is that regional cooperation must be strengthened to mitigate cross-border phishing risks effectively.

5.14.2 National-Level Gaps and Opportunities

Kenya's Data Protection Act (2019), Computer Misuse and Cybercrimes Act (CMCA, 2018), and National Cybersecurity Strategy (2014) provide a partial foundation, but they are misaligned with the risks posed by AI. The Data Protection Act lacks provisions on algorithmic transparency or AI-assisted decision-making. The CMCA criminalizes phishing but omits sandboxing, liability allocation, and audit mechanisms (Mwongera & Mbuthia, 2021). Fragmentation among the Office of the Data Protection Commissioner, the Communications Authority, and the Central Bank of Kenya weakens enforcement coherence (Omwansa & Waema, 2022). From an institutional theory

perspective, this fragmentation reduces coercive pressures, resulting in symbolic rather than substantive compliance.

The opportunity lies in adapting global best practices to the Kenyan context. As an illustration, implementing the EU AI Act in full would overwhelm MFIs, whereas a gradual approach to risk-tiered oversight and explainability requirements could achieve regulatory objectives while building institutional capacity.

5.14.3 Risk-Based and Adaptive Governance Lens

Under risk-based regulation, riskier applications, such as phishing detection, should be subject to more stringent regulation than low-risk tools, such as chatbots.

This is complemented by adaptive governance, which emphasizes experimental regulation and learning. Collectively, these theories offer a roadmap for Kenya: regulatory sandboxes should be introduced to enable supervised testing of AI defenses, phishing detection should be classified as a high-stakes risk, and oversight should be scaled as institutional capacity increases (Veale & Borgesius, 2021; Brown & Moller, 2022; Papagiannidis, 2025). This strategy is innovative and accountable, and it avoids premature overregulation.

5.14.4 Feasibility and Institutional Capacity

Ambitious reforms without institutional preparedness are bound to fail. Currently, agencies such as KE-CIRT/CC and CBK lack the experience and personnel to conduct algorithmic audits or certify AI models. It is thus necessary to adopt a gradual approach: start with sandboxes and explainability needs, build technical capacity via collaborations and donor assistance, and only then create a certification authority. The judiciary and the legislature will also need capacity

building to adjudicate algorithmic liability cases. In the absence of such incrementalism, reforms would merely be symbolic and continue to harbor the same weaknesses of the reform agenda.

5.15 Practical Interventions for MFIs

Although national and regional reforms are necessary to build an enabling environment, the strength of microfinance institutions (MFIs) against phishing also relies on institutional-level interventions. The results of the current study highlight several actionable measures that MFIs can implement within their existing resource constraints.

To start with, phishing awareness and response training should be institutionalized. High staff susceptibility persists, indicating not only low defenses but also organizational and behavioral gaps. Simulation-based training programs can be conducted regularly to decrease the click-through rates and establish a culture of vigilance (Alqaralleh et al., 2020). This is especially critical in MFIs, where technical weaknesses are usually exacerbated by human error. By considering the socio-technical aspects of vulnerability, training reinforces the idea that both technological and human resilience can be strengthened, making systems safer (Baashar et al., 2021).

Second, implement multi-layered security measures specifically designed to defend against phishing. The analysis has established that poor password habits and the absence of MFA are still prevalent. Some affordable yet highly impactful measures that MFIs can focus on include multi-factor authentication (MFA), domain-based message authentication, reporting, and conformance (DMARC), and formal incident response procedures. Such actions indicate a socio-technical equilibrium: as GRU-based detection systems advance adaptive defense, basic technical protection plays a significant role in minimizing successful breaches (Dwivedi et al., 2021).

Third, incorporate explainable AI (XAI) as a part of the phishing detection. Findings from validation indicated that although staff members rated GRU alerts as timely, there was reluctance to adopt them due to low transparency. Post-hoc explainability tools such as SHAP or LIME should be incorporated to help employees understand the features (e.g., suspicious URLs, abnormal sender history) that triggered an alert. This instills user trust, improves adoption, and aligns with new regulatory expectations regarding algorithmic transparency (Lundberg et al., 2023). In the case of MFIs, when levels of technical literacy differ, explainability is not a feature but a requirement for effective human-AI interaction.

Fourth, enhance co-operation through sectoral and regional platforms. MFIs tend to work in silos, which restricts access to threat intelligence and best practices. Having collective measures, either in the form of credit unions, national MFI associations, or regional organizations like the East African Community (EAC), may facilitate mutual learning on phishing trends, the sharing of phishing intelligence, and the implementation and deployment of AI. According to institutional theory, more pronounced normative pressures from such associations may promote substantive rather than symbolic adoption of key safeguards (Mutua, 2022). The cooperation lowers the costs and creates a shared defense stance against cross-border phishing.

Lastly, implement a gradual, layered approach to AI adoption. Resource-constrained MFIs ought to pilot GRU-based phishing detectors in a controlled setting and then gradually scale them. Pilots enable institutions to test usability, trust, and integration issues without excessive resource allocation. This adaptive governance will ensure that lessons learned during pilots are reflected in the phased rollouts, building institutional capacity and aligning with changing regulatory oversight (Papagiannidis, 2025).

Overall, these recommendations demonstrate that resilience in MFIs requires socio-technical balance characterized by human vigilance, layered defenses, explainable AI, institutional collaboration, and phase innovation. Incorporating these controls, MFIs can mitigate vulnerabilities revealed in this study, such as high vulnerability, ineffective protections, lack of trust in AI outputs, and risks of symbolic compliance, while positioning themselves at the center of integrating sustainable AI-based defenses.

5.16 Summary

This chapter has synthesized the study's findings, moving beyond descriptive results to interpret their theoretical, institutional, and policy significance. Three integrated insights emerge from the analysis.

First, phishing in MFIs is a socio-technical challenge. Vulnerabilities cannot be understood purely as technical failures; they emerge from the interaction of weak defenses, human error, organizational gaps and lack of AI governance. The phishing vulnerability–solution matrix illustrates that AI-driven defenses must be embedded in institutional practices and staff preparedness to be effective (Baashar et al., 2021; Abubakar & Bala, 2024).

Second, AI-driven defenses such as GRUs offer technical promise but face barriers to adoption. While GRUs achieved high detection accuracy, challenges of interpretability, linguistic bias, and limited validation highlight the importance of explainability, multilingual adaptability, and phased rollouts. These findings reinforce recent work showing that accuracy alone is insufficient for trustworthy AI in financial services (Lundberg et al., 2023; Dwivedi et al., 2021).

Third, governance frameworks remain misaligned with AI-driven risks. Kenya's CMCA (2018) and Data Protection Act (2019) provide partial coverage but omit algorithmic accountability, risk-

tiered oversight, and sandbox mechanisms. The six targeted reforms proposed in this study — including algorithmic audits and a national certification body — align with risk-based and adaptive governance principles, offering a roadmap for AI integration into financial cybersecurity (Veale & Borgesius, 2021; Brown & Möller, 2022; Papagiannidis, 2025).

Collectively, these insights demonstrate that safeguarding digital MFIs requires the co-evolution of AI innovation, governance reform, and institutional readiness. Technical solutions must be accurate, explainable, and adaptable; policies must balance ambition with feasibility; and institutions must build the capacity and culture needed to operationalize AI securely.

CHAPTER 6

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

6.1 Introduction

This chapter consolidates the study's insights into conclusions, contributions, and recommendations. It distills findings into cross-cutting themes that highlight their broader theoretical, institutional, and policy significance. It synthesizes significant conclusions that extend beyond individual objectives, identifies the study's key contributions, and translates them into actionable recommendations for microfinance institutions (MFIs), policymakers, and regulators.

By doing so, the chapter centrally positions the study within both current scholarly debates on artificial intelligence (AI) and cybersecurity and the practical realities of mitigating phishing in digital microfinance institutions. The aim is to demonstrate how AI-driven solutions, when aligned with governance reforms, institutional readiness, and human awareness, can strengthen resilience against evolving cyber threats.

6.2 Integrated Conclusions

Findings of this research indicate four general conclusions that transcend the specific objectives and demonstrate that phishing resilience among MFIs is an interdisciplinary phenomenon.

To begin with, phishing is a structural weakness of a socio-technical nature. Although technical vulnerabilities cannot be ignored, the persistence of vulnerabilities among MFI staff shows that phishing attacks exploit organizational and human factors as much as infrastructure weaknesses. This underscores that effective protection must account for human awareness, organizational culture, and technical security, in line with socio-technical systems theory (Baashar et al., 2021; Abubakar & Bala, 2024).

Second, the risks associated with AI are not aligned with governance and institutional frameworks. International standards such as ISO/IEC 27001 and the National Institute of Standards and Technology Cybersecurity Framework (NIST CSF) contain comprehensive general controls but lack AI-specific provisions. Data Protection Act (2019) and Computer Misuse and Cybercrimes Act (2018) of Kenya only partially address phishing. In particular, they lack algorithmic accountability, tiered oversight, and the use of sandboxes, which are gaining momentum in global discourse, including provisions grounded in the EU AI Act (Veale & Borgesius, 2021; Brown & Moller, 2022). This highlights the importance of localized and adaptive governance reforms that can align innovations with accountability.

Third, AI-based defenses may be implemented in resource-constrained MFIs, but their effectiveness depends on design choices. The GRU-based model developed in this research showed superior performance to classical classifiers while retaining its computational efficiency. This confirms that lightweight recurrent architectures can provide a satisfactory balance between accuracy and feasibility. It demonstrates that the success of AI utilization in MFIs is not determined solely by technical performance but also by the careful selection of architectures suited to the realities of resource-constrained MFIs (Mahmud et al., 2024; Sayeed & Dutta, 2025).

Fourth, adoption essentially relies on trust, explainability, and contextual adaptation. Pilot validation indicated that users valued timely notifications but remained hesitant due to a lack of transparency and their overreliance on English-only datasets. These findings underscore that it is impossible to achieve sustainability through accuracy alone. Explainable AI (XAI), support for multilingual environments, and adaptability to organizational workflows should be included in AI defenses to build confidence among users and regulators (Lundberg et al., 2023; Dwivedi et al., 2021).

Collectively, these findings show that phishing resilience among MFIs must be approached holistically: through technological innovation via AI, adaptive and risk-based governance frameworks, organizational preparedness, and long-term investment in people-centered strategies.

6.3 Contributions of the Study

This research makes remarkable contributions across theoretical, methodological, empirical, and policy fields.

This research makes significant contributions across the theoretical, methodological, empirical, and policy domains.

Within the theoretical domain, the study contributes to cybersecurity by framing phishing in Microfinance Institutions (MFIs) as a socio-technical vulnerability influenced by organizational practices, human behavior, technological gaps, and governance mechanisms. Building on the Socio-Technical Systems (STS) Theory, the study extends the classical framework by introducing a fourth dimension—the governance subsystem—which encompasses regulatory frameworks, AI-specific policies, algorithmic accountability, and compliance oversight. This enhanced STS framework proposes that organizational cybersecurity resilience is shaped by the interaction and alignment of four interdependent subsystems: human, organizational, technical, and governance. In addition, the study maps phishing threat typologies to AI-based defenses, broadening theoretical understanding of how recurrent neural networks, particularly GRUs, can be adapted to the operational realities of resource-constrained microfinance institutions.

Within the methodological domain, the study develops a replicable template for evaluating AI-driven phishing detection in resource-constrained environments. It moves beyond benchmarking by combining open phishing repositories with anonymized MFI-specific data and by validating

the model through sandbox testing, live deployment, and user feedback. This multi-stage design demonstrates how AI-based security research can be contextualized for practical environments (Mahmud et al., 2024).

In the empirical field, the research provides rare field-based evidence of AI deployment in Kenyan MFIs. While most prior studies remain confined to laboratory settings, this work demonstrates both feasibility and user trust in a real-world microfinance institution, addressing a key evidence gap in the literature (Abubakar & Bala, 2024).

At the governance level, the study recommends a sector-specific policy agenda for integrating AI-based solutions and guidelines into Kenya's cybersecurity framework. It identifies key gaps in the CMCA (2018). It proposes reforms such as regulatory sandboxes, certification mechanisms, and risk-tiered oversight, thereby adapting global principles (e.g., EU AI Act, MAS AI Principles) to Kenyan institutional contexts (Brown & Möller, 2022; Papagiannidis, 2025).

Together, these contributions enrich theory by embedding phishing resilience in socio-technical analysis, establishing a methodological and empirical blueprint for AI deployment in MFIs, and generating actionable policy insights for adaptive AI governance in emerging economies.

6.3.1 Contribution to Knowledge

This study contributed to knowledge by advancing the understanding of AI-driven phishing mitigation in digital microfinance institutions (MFIs) through theoretical, conceptual, methodological, and policy innovations. The novelty of the study lies not in proposing a new deep learning architecture, but in integrating established artificial intelligence techniques within a context-specific framework for enhancing cybersecurity resilience in resource-constrained financial institutions.

First, the study extended the Socio-Technical Systems (STS) Theory by explicitly incorporating a governance subsystem comprising regulatory frameworks, AI governance, compliance, and institutional oversight. Existing applications of STS Theory primarily focus on the interaction between human, organizational, and technical subsystems. By recognizing governance as an integral subsystem influencing cybersecurity resilience, the study provides an enhanced theoretical perspective for understanding AI adoption within digital MFIs.

Second, the study integrated Socio-Technical Systems Theory, Institutional Theory, and the Unified Theory of Acceptance and Use of Technology (UTAUT) into a single operational framework. This integration demonstrated how technological capability, organizational preparedness, user behavior, and regulatory compliance collectively influence the successful implementation of AI-driven phishing detection systems, thereby providing a coherent theoretical foundation for the study.

Third, the study developed a context-specific Phishing Vulnerability–Solution Matrix that systematically linked phishing vulnerabilities identified in digital MFIs with appropriate AI-driven mitigation strategies. Unlike conventional threat classifications, the matrix provides a practical decision-support framework for prioritizing AI interventions within resource-constrained financial institutions.

Finally, the study contributed methodological and policy knowledge by demonstrating a multi-stage framework for developing and validating AI-driven phishing detection models using publicly available datasets, institutional data, sandbox testing, and live pilot deployment, while also proposing context-specific amendments to Kenya's Computer Misuse and Cybercrimes Act (2018) to strengthen AI governance. Together, these contributions provide a practical and evidence-based

foundation for advancing AI-enabled cybersecurity resilience within digital microfinance institutions.

6.4 Practical and Policy Recommendations

The research findings highlight that phishing resilience requires institutional readiness and regulatory accountability.

At the institutional level, MFIs should progressively integrate lightweight AI models such as GRUs, adopting phased rollouts that begin with controlled pilots. These tools must be embedded within organizational practices that include phishing awareness programs, simulation-based training, and clear incident response protocols. Collaboration through national associations and regional platforms should be prioritized to enable the sharing of collective intelligence, reduce costs, and amplify resilience (Mutua, 2022).

At the regulatory level, supervisors should institutionalize regulatory sandboxes to support supervised experimentation and build regulatory expertise. A centralized certification mechanism, coordinated by KE-CIRT/CC and CBK, should establish standards for explainability, fairness, and robustness. Policymakers should further adopt a risk-based AI classification system in which phishing detection is designated as a high-risk application, warranting enhanced oversight. Sector-specific guidelines, developed by the Central Bank of Kenya (CBK) and the Capital Markets Authority (CMA), can provide clarity and guidance for financial institutions while ensuring alignment with both local realities and global best practices (Veale & Borgesius, 2021).

Ultimately, AI adoption in MFIs should be pursued as a dual process: strengthening institutional culture and capacity on one hand and embedding safeguards and accountability mechanisms through adaptive regulation on the other.

6.5 Recommended Amendments to the CMCA (2018)

Six specific reforms are recommended to update the CMCA and make it consistent with AI-driven risks. Based on the identified policy gaps, Table 5.1 outlines potential amendments to the Computer Misuse and Cybercrimes Act (2018) in Kenya, specifying the nature of each change and the respective entities involved. These recommendations were informed by the study findings, particularly the demonstrated performance and computational efficiency of the GRU model, as well as the governance, institutional, and resource constraints identified within the MFI context.

Table 6.1 Proposed Amendments to CMCA (2018)

Recommendation	**Description**	**Responsible Entity**
Regulatory Sandboxes	Pilot AI systems under supervised conditions	KE-CIRT/CC, CBK
Legal Clarity on AI Decisions	Define liability & require human oversight	Parliament, Judiciary
AI Certification Mechanism	National audit & certification body for AI	KE-CIRT/CC, KEBS
Algorithmic Transparency	Mandate explainability & disclosure standards	Parliament, Data Commissioner
Sector-Specific AI Guidelines	Finance-sector standards for AI in MFIs	CBK, CMA
AI Risk Classification	Tiered oversight (high/medium/low risk)	National AI Taskforce

Source: Author (2025).

The proposed reforms combine proactive tools, such as sandboxes and certification programs, with accountability tools, such as transparency and liability controls. Together, these amendments would transform CMCA (2018) into a more adaptable and risk-responsive framework in Kenya, more consistent with international best practices and the realities in which MFIs operate. These amendments localize international best practices to Kenyan realities, making reforms more feasible for MFIs and less idealistic.

Implementation could follow a phased approach, beginning with controlled pilots and progressing to wider institutional adoption after evaluating performance, operational requirements, and user acceptance. The main costs would relate to infrastructure, training, data management, model maintenance, and cybersecurity monitoring. The GRU model's balance between detection performance and computational efficiency could support adoption by resource-constrained MFIs, although actual costs would vary across institutions.

Sustainable adoption would require continuous model evaluation and retraining, staff awareness, management support, and appropriate governance. MFIs with limited technical capacity could also consider shared or managed cybersecurity services. The framework further presents potential for commercialization as a managed phishing-detection service, although its commercial viability would require further assessment of cost, scalability, data protection, and regulatory requirements.

At the policy level, regulatory sandboxes, certification, transparency requirements, and risk classification could provide a practical mechanism for testing and regulating AI-driven cybersecurity solutions before wider adoption.

6.6 Future Research Directions

Like any focused study, this research provides a foundation for further scholarly exploration shaped by contextual and methodological factors, pointing to valuable areas for future research.

The initial area for future research stems from the linguistic and cultural bias in the information, which included English-language phishing messages and SMS. Although this plays a key role in training baselines, it overlooks the fact that phishing in Kenya and other African regions often involves code-switching between Kiswahili and local languages. This bias diminishes detection accuracy in multilingual cases. Future research should focus on developing multilingual phishing datasets that incorporate local linguistic and cultural features to improve detection accuracy across diverse populations and regions (Nguyen et al., 2020; Ouedraogo et al., 2023).

The second area for further research concerns the limited institutional coverage of validation. Although the pilot implementation in a Kenyan MFI yielded valuable field data, the heterogeneity of Tier II and Tier III institutions, their resource profiles, and their customer bases remains to be studied. By increasing the number of pilots across numerous MFIs and markets, external validity would be enhanced, and the generalizability of findings would increase in other organizational and regulatory contexts (Heckmann et al., 2022; Abubakar & Bala, 2024).

The third aspect is that GRU models behave like “black boxes”, limiting interpretability and user trust. Post hoc explainability systems, such as SHAP and LIME, are suggested, but they remain inadequate for accountability in high-stakes financial ecosystems. Further studies are needed to investigate the native explainability embedded in model architectures, including hybrid models such as GRU-transformer models, to balance performance and explainability (Lundberg et al., 2023).

The fourth area for future research is associated with the dynamism of phishing threats. Simulated attacks and static datasets cannot adequately represent the flexibility of real attackers. In the future, research should explore continuous learning architecture development to enable near-real-time updates, supported by intelligence-sharing structures among MFIs (Papagiannidis, 2025).

Lastly, future research should also examine non-technical barriers that influence adoption and scalability. Institutional resistance, budgetary constraints, and regulatory uncertainty usually hinder the operationalization of AI-based defenses in the African MFIs. Besides, the policy analysis was context-specific to Kenya. Comparing the jurisdictional regulation of AI across the EU, Singapore, and other African nations may offer insights into both harmonized and localized modes of AI regulation (Mutua, 2022; Mbarika et al., 2021; Papagiannidis, 2025).

These priorities, such as multi-institutional validation, expanded multilingual datasets, interpretable model structures, adaptive learning systems, and comparative governance analyses, constitute a research agenda going forward to reinforce AI-based phishing mitigation in digital financial ecosystems.

REFERENCES

Abouraia, M. K. (2024). Artificial intelligence in finance: Navigating opportunities and risks. In Risks and challenges of AI-driven finance: Bias, ethics, and security (pp. 17–34). IGI Global.

Abubakar, A., & Bala, K. (2024). Artificial intelligence adoption in African financial institutions: Opportunities, challenges, and regulatory pathways. Journal of Financial Innovation in Emerging Markets, 6(2), 45–62.

Adebowale, M. A., Lwin, K. T., & Hossain, M. A. (2023). Intelligent phishing detection scheme using deep learning algorithms. Journal of Enterprise Information Management, 36(3), 747–766.

Adnan, S. A., & Kumar, P. (2021). Role of microfinance in economic development. Adhyayan: A Journal of Management Sciences, 11(2), 22–30.

Afzal, M., Meraj, M., & Kaur, M. (2024). How does cybersecurity awareness help achieve digital financial inclusion in rural India amid an escalating cyber fraud scenario? Journal of Cyber Security Technology.

Ahmad, I., Iqbal, S., Jamil, S., & Kamran, M. (2021). A systematic literature review of e-banking frauds: Current scenario and security techniques. Linguistica Antverpiensia, 2(2), 3509–3517.

Åkerblad, L., Seppänen-Järvelä, R., & Haapakoski, K. (2021). Integrative strategies in mixed methods research. Journal of Mixed Methods Research, 15(2), 152–170.

Alam, M. S. (2024). A study of digital transformation in the microfinance landscape of Bangladesh.

Alanezi, M. (2021). Phishing detection methods: A review.

Ali, G., Mijwil, M. M., Buruga, B. A., & Abotaleb, M. (2024). A comprehensive review of cybersecurity issues and their mitigation measures in FinTech.

Ali, M., Rahman, H., & Hussain, S. (2023). Lightweight recurrent models for phishing detection in low-resource settings. IEEE Access, 11, 23745–23759.

Aljofey, A., Jiang, Q., Qu, Q., Huang, M., & Niyigena, J. P. (2020). An effective phishing detection model based on character level convolutional neural network from URL. Electronics, 9(9), 1514.

Alkhalil, Z., Hewage, C., Nawaf, L., & Khan, I. (2021). Phishing attacks: A recent comprehensive study and a new anatomy. Frontiers in Computer Science, 3, 563060.

Almiani, M., Alauthman, M., AlZain, M. A., Al-Duwairi, B., & Al-Utaibi, K. A. (2023). Deep learning-based phishing detection: A comparative study of evaluation metrics. Journal of Information Security and Applications, 71, 103419.

Almujahid, N. F., Haq, M. A., & Alshehri, M. (2024). Comparative evaluation of machine learning algorithms for phishing site detection. PeerJ Computer Science, 10, e2131.

Alshingiti, A., Khan, I., & Noor, M. (2023). Efficiency-driven deep neural network models for phishing threat analysis. Computers & Security, 130, 103307.

Alshingiti, Z., Alaqel, R., Al-Muhtadi, J., Haq, Q. E. U., Saleem, K., & Faheem, M. H. (2023). A deep learning-based phishing detection system using CNN, LSTM, and LSTM-CNN. Electronics, 12(1), 232.

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. Journal of Big Data, 8, 1–74.

Anti-Phishing Working Group. (2024). Phishing activity trends report. https://apwg.org/trendsreports/

Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., ... & Herrera, F. (2020). Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Ayres, I., & Braithwaite, J. (2023). Responsive regulation: Transcending the deregulation debate. Oxford University Press. https://doi.org/10.1093/oso/9780195070705.001.0001

Baashar, Y., Alharthi, H., Alkawsi, G., & Alzahrani, A. (2021). A socio-technical perspective on cybersecurity resilience in digital financial ecosystems. Journal of Cyber Policy, 6(1), 89–108.

Babu Nuthalapati, S. (2023). AI-enhanced detection and mitigation of cybersecurity threats in digital banking. Educational Administration: Theory and Practice, 29(1), 357–368.

Bardhan, A. K., Nag, B., Mishra, C. S., & Tarei, P. K. (2021). An integrated framework for analysing performance indicators of Indian microfinance institutions: A multi-stakeholder perspective. Benchmarking: An International Journal, 28(9), 2711–2740.

Bell, E., & Bryman, A. (2023). Sampling in qualitative research. In Business research methods. Oxford University Press. https://doi.org/10.1093/hebz/9780198869443.003.0030

Braun, V., & Clarke, V. (2023). Toward good practice in thematic analysis: Avoiding common problems and be(com)ing a knowing researcher. International Journal of Transgender Health, 24(1), 1–6. https://doi.org/10.1080/26895269.2022.2129597

Brown, I., & Möller, N. (2022). Regulating artificial intelligence in financial services: Risk-based approaches and adaptive governance. AI and Society, 37(3), 987–1002. https://doi.org/10.1007/s00146-021-01191-8

Camacho, J. (2024). Privacy-preserving AI for financial cybersecurity: Federated approaches to phishing detection. Information Systems Frontiers, 26(2), 425–439.

Camacho, N. G. (2024). The role of AI in cybersecurity: Addressing threats in the digital age. Journal of Artificial Intelligence General Science (JAIGS), 3(1), 143–154.

Canadian Institute for Cybersecurity. (2017). CICIDS 2017 dataset. University of New Brunswick. https://www.unb.ca/cic/datasets/ids-2017.html

Carnegie Mellon University. (2004). Enron email dataset. School of Computer Science. https://www.cs.cmu.edu/~enron/

Catal, C., Giray, G., Tekinerdogan, B., Kumar, S., & Shukla, S. (2022). Applications of deep learning for phishing detection: A systematic literature review. Knowledge and Information Systems, 64(6), 1457–1500.

Chaokromthong, K., & Sintao, N. (2021). Sample size estimation using Yamane and Cochran and Krejcie and Morgan and Green formulas and Cohen statistical power analysis by G*Power and comparisons. APHEIT International Journal of Interdisciplinary Social Sciences and Technology, 10(2), 76–86.

Chen, R., Zhang, L., & Wu, Y. (2023). Hybrid recurrent-attention networks for email phishing detection. Expert Systems with Applications, 227, 120307.

Colombari, R., Neirotti, P., & Berbegal-Mirabent, J. (2024). Disentangling the socio-technical impacts of digitalization: What changes for shop-floor decision-makers? International Journal of Production Economics, 276, 109377. https://doi.org/10.1016/j.ijpe.2024.109377

Communications Authority of Kenya. (2024). Cybersecurity report. https://www.ca.go.ke

DiMaggio, P. J., & Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 48(2), 147–160. https://doi.org/10.2307/2095101

Duarte, E., Dias, M., & Martins, A. (2021). Social engineering attacks in financial institutions: Emerging trends and mitigation strategies. Computers & Security, 105, 102239.

Dwivedi, Y. K., et al. (2021). Responsible AI: Integrating explainability into socio-technical systems. International Journal of Information Management, 60, 102367.

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994.

Elder-Vass, D. (2022). Pragmatism, critical realism and the study of value. Journal of Critical Realism, 21(3), 261–287.

Elder-Vass, D. (2022). Reconstructing realism: New perspectives in social ontology. Routledge.

EURAM. (2025). New institutional theory and AI: Algorithms as a driving force of organizational isomorphism. European Academy of Management Conference Proceedings.

European Commission. (2021). Proposal for a Regulation on Artificial Intelligence (Artificial Intelligence Act).

Fahmie, T. A., Rodriguez, N. M., Luczynski, K. C., Rahaman, J. A., Charles, B. M., & Zangrillo, N. A. (2023). Toward an explicit technology of ecological validity. Journal of Applied Behavior Analysis, 56(2), 302–322. https://doi.org/10.1002/jaba.972

Gal, M. S., & Aviv, O. (2020). The competitive effects of the GDPR. Journal of Competition Law & Economics, 16(3), 349–391.

Gandotra, E., & Gupta, D. (2021). An efficient approach for phishing detection using machine learning. In Multimedia security: Algorithm development, analysis and applications (pp. 239–253).

Gautam, N. (2021). Applications of AI and machine learning in banking solutions. Grenze International Journal of Engineering & Technology, 7(1).

Geetha, R., & Thilagam, T. (2021). A review on the effectiveness of machine learning and deep learning algorithms for cyber security. Archives of Computational Methods in Engineering, 28(4), 2861–2879.

Géron, A. (2022). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (3rd ed.). O'Reilly Media.

Gonçalves, M. E. (2020). The risk-based approach under the new EU data protection regulation: A critical perspective. Journal of Risk Research, 23(2), 139–152.

Hadfield, G. (2023). Legal sandboxes and the future of regulation. SSRN.

Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2021). A primer on partial least squares structural equation modeling (PLS-SEM) (3rd ed.). Sage Publications. https://doi.org/10.1007/978-3-030-80519-7

Heckmann, C., Wendt, T., & Klein, M. (2022). AI adoption in financial organizations: Challenges of validation, scalability, and regulatory compliance. Journal of Banking Regulation, 23(2), 145–160.

Hillman, D., Harel, Y., & Toch, E. (2023). Evaluating organizational phishing awareness training on an enterprise scale. Computers & Security, 132, 103364.

Hochreiter, S. (2020). Recurrent networks and the long-term dependency problem revisited. Neural Computation, 32(8), 1581–1600.

Hoong, Y., Rezania, D., & Baker, R. (2024). When traditional SME managers encounter cybersecurity: Discourse analysis of opportunities and dilemmas in meeting the demands. Technology in Society, 78, 102650. https://doi.org/10.1016/j.techsoc.2024.102650

Ibraheem, I. O., & Tosho, A. U. Zero day attack vulnerabilities: Mitigation using machine learning for performance evaluation. Journal of Computers for Society, 5(1), 43–58.

Ibrahimnur, A. A. (2023). Impact of cybercrime on the finance sector: A case of banks in Nairobi County, Kenya (2008–2022) (Doctoral dissertation, University of Nairobi).

ICIS. (2023). Institutional responses to AI governance pressures in digital transformation. Proceedings of the International Conference on Information Systems (ICIS 2023). https://aisel.aisnet.org/icis2023

International Criminal Police Organization. (2024). African cyberthreat assessment report 2024. https://www.interpol.int/en/Crimes/Cybercrime/Projects/AFJOC-African-Joint-Operation-against-Cybercrime

International Monetary Fund. (2023). The rise and impact of fintech in Latin America. Fintech Notes, 2023(003). https://doi.org/10.5089/9798400235474.063

International Monetary Fund. (2024). Global financial stability report, April 2024: The last mile—Financial vulnerabilities and risks. https://doi.org/10.5089/9798400257704.082

Islam, M. R., Abawajy, J., & Din, S. (2022). Phishing email detection using machine learning and hybrid models. Computers & Security, 116, 102673.

Javaheri, D., Fahmideh, M., Chizari, H., Lalbakhsh, P., & Hur, J. (2023). Cybersecurity threats in FinTech: A systematic review. Expert Systems with Applications, 122697.

Johnson, D. (2023). Responsive regulation in practice: A review of the international academic literature. Regulation & Governance, 17(4), 1201–1215. https://doi.org/10.1111/rego.12487

Johnson, D. R. (2021). Governing AI in the global south: Regulatory asymmetries and implementation gaps. Global Policy, 12(4), 510–521. https://doi.org/10.1111/1758-5899.12947

Johnson, R. B., & Christensen, L. B. (2024). Educational research: Quantitative, qualitative, and mixed approaches. Sage Publications.

Joseph, J. (2023). The effectiveness of cybersecurity measures in preventing cyber-attacks on mobile payment systems: A case study of Moshi District.

Kaloudi, N., & Li, J. (2020). The AI-based cyber threat landscape: A survey. ACM Computing Surveys, 53(1), 1–34.

Kamiri, J., & Mariga, G. (2021). Research methods in machine learning: A content analysis. International Journal of Computer and Information Technology, 10(2), 78–91.

Kar, D., Saxena, K., & Wary, A. (2024, August). Comparative study of the performance of GRU and BERT in phishing URL detection. In Proceedings of the International Conference on Innovative Computing & Communication (ICICC 2024).

Kaur, R., Gabrijelčič, D., & Klobučar, T. (2023). Artificial intelligence for cybersecurity: Literature review and future research directions. Information Fusion, 97, 101804. https://doi.org/10.1016/j.inffus.2023.101804

Kay, A., Hutcherson, C., Keene, C., Zhang, X., & Terwilliger, M. G. (2021). How financial institutions address cybersecurity threats: A critical analysis. Issues in Information Systems, 22(1), 63–74.

Kenya National Bureau of Statistics. (2019a). 2019 Kenya population and housing census. KNBS.

Kenya National Bureau of Statistics. (2019c). Summary report on Kenya's population projections – 2019 KPHC. KNBS.

Khaleefah, A. D., & Al-Mashhadi, H. M. (2024). Methodologies, requirements, and challenges of cybersecurity frameworks: A review. Iraqi Journal of Science.

Khatri, K. K. (2020). Research paradigm: A philosophy of educational research. International Journal of English Literature and Social Sciences, 5(5), 1435–1440.

Khekare, G., Bramhane, L. K., Dhule, C., Agrawal, R., & Turukmane, A. V. (2022). Testing and analysis of predictive capabilities of machine learning algorithms. In Integrating meta-heuristics and machine learning for real-world optimization problems (pp. 419–442). Springer.

Kiganda, M. (2022). An assessment of the factors affecting cyber resilience in microfinance institutions in Kenya (Doctoral dissertation, Strathmore University).

Kogen, L. (2024). Qualitative thematic analysis of transcripts in social change research: Reflections on common misconceptions and recommendations for reporting results. International Journal of Qualitative Methods, 23. https://doi.org/10.1177/16094069231225919

Lawani, A. (2021). Critical realism: What you should know and how to apply it. Qualitative Research Journal, 21(3), 320–333.

Leavy, P. (2022). Research design: Quantitative, qualitative, mixed methods, arts-based, and community-based participatory research approaches. Guilford Publications.

Li, Y., Xu, J., & Zhao, X. (2024). Transformer-based semantic models for advanced phishing email detection. Information Sciences, 665, 119945.

Lundberg, S. M., & Lee, S. I. (2023). Interpretable machine learning for high-stakes domains: Advances in SHAP and LIME for financial services. Patterns, 4(7), 100734.

Lundberg, S. M., et al. (2023). Explainable AI and model interpretability for responsible machine learning. Nature Machine Intelligence, 5(3), 210–220.

Mahmud, K., Rahman, M., & Khan, S. (2024). Lightweight deep learning models for cybersecurity: A case for GRU architectures in resource-constrained environments. Computers & Security, 135, 103469.

Malatji, M. (2023). Management of enterprise cyber security: A review of ISO/IEC 27001:2022. In 2023 International Conference on Cyber Management and Engineering (CyMaEn) (pp. 117–122). IEEE.

Mallick, M. A. I., & Nath, R. (2024). Navigating the cybersecurity landscape: A comprehensive review of cyber-attacks, emerging trends, and recent developments. World Scientific News, 190(1), 1–69.

Mbarika, V., Mutua, D., & Kiarie, J. (2021). Institutional adoption of ICT frameworks in Sub-Saharan Africa: Contextual limitations and adaptive models. African Journal of Information Systems, 13(2), 45–63.

Mbarika, V., Samoilenko, S., & Boateng, R. (2021). Institutional constraints in technology adoption: Lessons for AI-driven financial services in Africa. Information Technology for Development, 27(4), 639–656.

Mingers, J., Mutch, A., & Willcocks, L. (2020). Critical realism in information systems research. MIS Quarterly, 44(4), 1715–1738.

Momand, A., Jan, S. U., & Ramzan, N. (2023). A systematic and comprehensive survey of recent advances in intrusion detection systems using machine learning: Deep learning, datasets, and attack taxonomy. Journal of Sensors, 2023, Article 6048087. https://doi.org/10.1155/2023/6048087

Momand, A., Jan, S. U., & Ramzan, N. (2023). A systematic and comprehensive survey of recent advances in intrusion detection systems using machine learning: Deep learning, datasets, and attack taxonomy. Journal of Sensors, 2023, Article 6048087. https://doi.org/10.1155/2023/6048087

Monetary Authority of Singapore. (2021). Principles to promote fairness, ethics, accountability and transparency (FEAT). https://www.mas.gov.sg/publications

Morris, J., Tatschner, S., Heinl, M. P., Heinl, P., Newe, T., & Plaga, S. (2023). Cybersecurity as a service. In Cybersecurity vigilance and security engineering of Internet of Everything (pp. 141–161). Springer Nature Switzerland.

Mutai, P. (2022). Cybersecurity governance in Africa: Implementation challenges of the AU Malabo Convention. African Journal of Information and Communication, 29, 1–20.

Mutua, D. (2022). Institutional readiness for AI governance in financial services. Journal of African Digital Finance, 4(1), 56–71.

Myers, M. D., & Klein, H. K. (2021). Epistemological pluralism in information systems research. Information Systems Journal, 31(2), 223–239.

Nadeem, M., Zahra, S. W., Abbasi, M. N., Arshad, A., Riaz, S., & Ahmed, W. (2023). Phishing attack, its detections and prevention techniques. International Journal of Wireless Security and Networks, 1(2), 13–25.

Nairobi City County Government. (2023). Nairobi City County land use policy, 2023. Nairobi City County Government.

Nature. (2024). Governing artificial intelligence: Between regulatory inertia and innovation. Humanities and Social Sciences Communications, 11(1), 451.

Nguyen, T., Pham, H., & Le, D. (2020). Building multilingual phishing corpora for natural language processing applications. Journal of Information Security and Applications, 54, 102563. https://doi.org/10.1016/j.jisa.2020.102563

Nicholls, J., Kuppa, A., & Le-Khac, N. A. (2021). Financial cybercrime: A comprehensive survey of deep learning approaches to tackle the evolving financial crime landscape. IEEE Access, 9, 163965–163986.

Noh, S. H. (2021). Analysis of gradient vanishing of RNNs and performance comparison. Information, 12(11), 442.

Non, R. (2023). Exploring the challenges faced by microfinance industry in facilitating ecosystem financing for microenterprises in Malaysia (Doctoral dissertation, University of Wales Trinity Saint David).

Nwankwo, C., Kanyangale, M., Anoke, A. F., & Eze, S. U. (2023). Effect of cyber security on business sustainability of listed microfinance banks in Nigeria. Artha Journal of Social Sciences, 22(1), 79–106.

Olabiyi, O. (2024). Evaluating traditional versus AI-based phishing detection models. Journal of Cybersecurity Research, 6(1), 45–63.

Oner, U., Cetin, O., & Savas, E. (2025). Human factors in phishing: Understanding susceptibility and resilience. Computer Standards & Interfaces, 94, 104014.

Ouedraogo, A., Coulibaly, K., & Traore, F. (2023). Multilingual AI for cybersecurity: Addressing language bias in African digital finance. ACM Transactions on Asian and Low-Resource Language Information Processing, 22(4), 1–18.

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. The BMJ, 372, n71. https://doi.org/10.1136/bmj.n71

Papagiannidis, S. (2025). Adaptive AI regulation for financial institutions: Lessons from emerging economies. Government Information Quarterly, 42(1), 101–112.

Papagiannidis, S. (2025). Adaptive governance for AI in financial ecosystems: Lessons from sandboxing and iterative regulation. Government Information Quarterly, 42(1), 101791.

Papagiannidis, S. (2025). Responsible AI governance: Conceptual foundations and organizational practices. Technological Forecasting & Social Change, 208, 122498. https://doi.org/10.1016/j.techfore.2024.122498

Paraschiv, D., Toader, L., Nițu, M., & Negrea, Ș. (2021, June). Internet fraud and phishing attacks—A European perspective. In 7th BASIQ International Conference on New Trends in Sustainable Business and Consumption (pp. 3–5).

Pattnaik, D., Ray, S., & Raman, R. (2024). Applications of artificial intelligence and machine learning in the financial services industry: A bibliometric review. Heliyon.

Paul, R. (2023). European artificial intelligence "trusted throughout the world": Risk-based regulation and the fashioning of a competitive common AI market. Regulation & Governance. https://doi.org/10.1111/rego.12563

Qu, Z., Su, L., Wang, X., Zheng, S., Song, X., & Song, X. (2018, January). A unsupervised learning method of anomaly detection using GRU. In 2018 IEEE International Conference on Big Data and Smart Computing (BigComp) (pp. 685–688). IEEE.

Rieder, A., Chakraborty, S., Goyal, S., & Berndt, D. J. (2025). A critical realist approach to agent-based modeling: Unlocking prediction in non-positivist paradigms. Information Systems Journal, 40(2). https://doi.org/10.1177/02683962241280657

Rizvi, S., Akhtar, M., & Khan, F. (2022). Integration challenges of AI-based models in financial cybersecurity systems. Journal of Digital Finance, 9(2), 211–230.

Roy, A., Chatterjee, S., & Paul, P. (2022). GRU-driven real-time phishing detection: Balancing accuracy and resource efficiency. Journal of Information Security and Applications, 68, 103238.

Roy, S. S., Awad, A. I., Amare, L. A., Erkihun, M. T., & Anas, M. (2022). Multimodel phishing URL detection using LSTM, bidirectional LSTM, and GRU models. Future Internet, 14(11), 340.

Rwashdeh, B., Awad, A., & Butt, U. (2025). Enhancing the incident response capabilities of the UAE critical infrastructure against phishing cyber security threat: An AI-based approach. In K. Al Marri et al. (Eds.), BUiD Doctoral Research Conference 2024 (pp. 193–205). Springer. https://doi.org/10.1007/978-3-031-84371-6_16

Salem, F. M. (2021). Gated RNN: The gated recurrent unit (GRU) RNN. In Recurrent neural networks: From simple to gated architectures (pp. 85–100). Springer International Publishing.

Sayeed, M., & Dutta, S. (2025). Recurrent neural networks for phishing detection: A comparative study of GRU, LSTM, and hybrid models. IEEE Transactions on Information Forensics and Security, 20(3), 456–468. https://doi.org/10.1109/TIFS.2024.3354782

Schoch, K. (2020). Case study research. In Research design and methods: An applied guide for the scholar-practitioner (Vol. 31, Issue 1, pp. 245–258).

Schuster, M., & Paliwal, K. (2020). Bidirectional recurrent networks for contextual text analysis. IEEE Transactions on Signal Processing, 68, 2463–2475.

Shafiq, M., Ahmad, I., & Kim, S. (2020). Collaborative intelligence in phishing prevention for microfinance systems. Computers & Security, 95, 101874.

Shah, V. (2021). Machine learning algorithms for cybersecurity: Detecting and preventing threats. Revista Espanola de Documentacion Cientifica, 15(4), 42–66.

Somesha, M., Pais, A. R., Rao, R. S., & Rathour, V. S. (2020). Efficient deep learning techniques for the detection of phishing websites. Sādhanā, 45, 1–18.

Somesha, S., Njoroge, M., & Wafula, J. (2020). AI adaptation barriers in microfinance cybersecurity. Journal of Financial Technology, 4(3), 175–190.

Taofeek, A. O. (2024). Development of a novel approach to phishing detection using machine learning. ATBU Journal of Science, Technology and Education, 12(2), 336–351.

Tay, Y., Bahri, D., Metzler, D., Juan, D. C., Zhao, Z., & Zheng, C. (2021). Synthesizer: Rethinking self-attention for transformer models. In International Conference on Machine Learning (pp. 10183–10192). PMLR.

Tech For Good Institute. (2023). Towards a resilient cyberspace in Southeast Asia. https://techforgoodinstitute.org/research/tfgi-reports/towards-a-resilient-cyberspace-in-southeast-asia/

van Baalen, S., & Boon, M. (2024). Understanding disciplinary perspectives: A framework to develop skills for interdisciplinary research collaborations of medical experts and engineers. BMC Medical Education, 24, Article 1000. https://doi.org/10.1186/s12909-024-05913-1

Veale, M., & Borgesius, F. Z. (2021). Demystifying the EU's AI Act. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402

Venkatesh, V., Thong, J. Y. L., & Xu, X. (2022). Unified theory of acceptance and use of technology (UTAUT): 2022 update and extensions. MIS Quarterly, 46(3), 123–146.

Verma, R., & Das, A. (2022). Phishing Website Dataset (Version 1). Mendeley Data. https://data.mendeley.com/datasets/h3cgnj8hft/1

Webb, H. C., & Sallos, M. (2024). Navigating cybersecurity challenges in FinTech: Strategic approaches for SMEs in the GCC region. In The Palgrave handbook of FinTech in Africa and Middle East: Connecting the dots of a rapidly emerging ecosystem (pp. 1–19). Springer Nature Singapore.

Xu, L., Zhang, Y., & Liu, F. (2020). Deep learning-based email phishing detection: A comparative study. Computers & Security, 95, 101864.

Yamane, T. (1967). Statistics: An introductory analysis (2nd ed.). Harper & Row.

Zachariadis, M., Scott, S. V., & Barrett, M. (2022). Methodological implications of critical realism for digital transformation research. Journal of Strategic Information Systems, 31(3), 101727.

Zhang, Y., Li, X., Wang, H., & Chen, J. (2023). Artificial intelligence-enabled phishing attacks: Trends, challenges and defensive strategies. Journal of Cyber Security Technology, 7(2), 98–117.

APPENDICES

Appendix 1: Questionnaire

MICROFINANCE INSTITUTIONS PHISHING VULNERABILITY SURVEY

Declaration:

I am a PhD student at the School of Pure and Applied Sciences, Kirinyaga University, conducting research on adaptive AI-driven phishing detection frameworks in digital microfinance institutions. The study addresses issues related to phishing threats, cybersecurity vulnerabilities, and organizational readiness to adopt artificial intelligence (AI)-based detection systems in microfinance institutions (MFIs) operating within Nairobi.

I appreciate your participation in this survey. Your experiences, insights, and opinions regarding phishing incidents, cybersecurity practices, and institutional responses are valuable to this research. Participation in this study is completely voluntary and confidential. You will not be personally identified in any part of the report, and all responses will be anonymized. The information collected will be used solely for academic purposes in accordance with the approved research objectives and will not be disclosed or used for any non-research activities.

By completing this form, you consent to the use of your responses for research purposes only.

☐ I understand the purpose of this questionnaire and consent to participate

SECTION A: RESPONDENT & ORGANIZATIONAL PROFILE

1. **What is your current position or role within the organization?**

   ☐ Chief Executive Officer / Managing Director

   ☐ Head of Information Technology / ICT Manager

☐ Cybersecurity Specialist / Network Administrator

☐ Operations Officer / Frontline Service Staff

☐ Internal Auditor / Compliance Officer

☐ Other (please specify): _____

2. **How many years of professional experience do you have in the financial services sector?**

   ☐ Less than 1 year

   ☐ 1 to 3 years

   ☐ 4 to 6 years

   ☐ More than 6 years

3. **How many employees does your organization currently have?**

   ☐ Fewer than 10 employees

   ☐ 10 to 20 employees

   ☐ 21 to 50 employees

   ☐ 51 to 100 employees

   ☐ More than 100 employees

4. **Approximately how many clients does your institution serve?**

   ☐ Fewer than 500 clients

   ☐ 500 to 1,000 clients

   ☐ 1,001 to 5,000 clients

   ☐ More than 5,000 clients

5. **How would you describe your organization's current level of digital technology maturity?**

   ☐ Basic: Primarily manual processes, limited use of digital tools

   ☐ Intermediate: Some digital systems, such as mobile or core banking platforms, are in place

☐ Advanced: Use of APIs, automated platforms, AI-based tools, and integrated digital services

6. **Which of the following best describes your organization's primary geographical reach?**

   ☐ Operates within a single local community or branch

   ☐ Operates across multiple counties within Kenya

   ☐ Operates nationally across the entire country

   ☐ Operates across more than one country (cross-border/international)

SECTION B: CYBERSECURITY GOVERNANCE AND FRAMEWORK ADOPTION

7. **Does your organization currently follow a formal cybersecurity policy or framework?**

   ☐ Yes

   ☐ No

8. **If YES, which cybersecurity framework(s) does your organization use?**

   ☐ ISO/IEC 27001 (Information Security Management System)

   ☐ NIST Cybersecurity Framework

   ☐ COBIT

   ☐ A custom-developed internal policy /locally developed or in-house framework

   ☐ Other (please specify): _____

9. **How frequently are your organization's cybersecurity policies and protocols reviewed or updated?**

   ☐ At least once every 6 months

   ☐ After 1 year

   ☐ Every 2 to 3 years

☐ Irregularly (only after an incident)

☐ Policies are rarely or never updated

10. **What approximate percentage of your organization's annual operating budget is allocated to cybersecurity tools, training, and resources?**

    ☐ Less than 1%

    ☐ 1% to 2%

    ☐ 3% to 5%

    ☐ More than 5%

    ☐ The organization does not track this separately

11. **Does your organization conduct cybersecurity vulnerability assessments or penetration testing?**

    ☐ Yes, regularly (e.g., quarterly or annually)

    ☐ Yes, occasionally, but not on a fixed schedule

    ☐ No, these assessments are not conducted

    ☐ I do not know

12. **Is there a documented <u>Incident Response Plan</u> in place for cybersecurity events (e.g., phishing attacks, breaches)?**

    ☐ Yes, a formal and regularly tested plan exists

    ☐ A draft or informal plan exists but is not yet fully implemented

    ☐ No incident response plan currently exists

    ☐ I am not sure

13. **How confident are you in your knowledge of your organization's cybersecurity framework and practices?**

    ☐ Very confident

    ☐ Moderately confident

    ☐ Not confident

<u>SECTION C: TECHNICAL SECURITY CONTROLS AND INCIDENT HISTORY</u>

14. **Which of the following technical safeguards are currently implemented in your organization?** (Check all that apply)

    ☐ Antivirus and antimalware protection

    ☐ Firewall systems

    ☐ Intrusion Detection and Prevention Systems (IDS/IPS)

    ☐ Multi-Factor Authentication (MFA) for user accounts

    ☐ Virtual Private Network (VPN) usage

    ☐ Email filtering or anti-spam systems

    ☐ Patch management system for regular software updates

    ☐ Regular internal or third-party security audits

15. **Has your organization experienced one or more phishing incidents in the past 12 months?**

    ☐ Yes

    ☐ No

    ☐ I am not sure

16. **Which types of phishing or social engineering attacks were encountered?** (Check all that apply)

    ☐ Email spoofing or impersonation

    ☐ Malicious attachments or links in emails

    ☐ SMS or WhatsApp phishing (smishing)

    ☐ Fake websites or login pages

    ☐ Social media impersonation of staff or institution

    ☐ Other: _____

17. **How did employees typically respond to phishing attempts (based on your knowledge)?** (Check all that apply)

    ☐ Reported the email or incident to IT/security personnel

    ☐ Opened the email and clicked on a suspicious link

    ☐ Ignored the email but did not report it

☐ Used personal judgment to verify the email before responding

☐ I am not sure how the staff responded

18. **What do you think could have been the key vulnerabilities or weaknesses that allowed phishing to succeed?** (Check all that apply)

☐ Weak or reused passwords

☐ Lack of email authentication mechanisms

☐ Staff not trained in identifying phishing

☐ No two-factor authentication in place

☐ Software systems not updated or patched

☐ Lack of real-time monitoring or alerts

19. On average, how many phishing attempts emails does your organization detect per month (email, SMS, portal alerts)?

☐ None

☐ 1–5 attempts

☐ 6–10 attempts

☐ 11 - 20 attempts

☐ Over 20 attempts

☐ Not sure

20. How has your organization been impacted by phishing attacks? (Tick where appropriate)

		1. No impact	2. Minor impact	3. Moderate impact	4. Major impact	5. Severe impact	Please estimate the loss
C1.	Financial Loss						
C2.	Operational Delays						

C3.	Reputational damage						
C4.	Compliance Issues						

SECTION D: HUMAN & ORGANIZATIONAL SECURITY POSTURE

21. Please rate the following statements using a scale from 1 to 5. Tick **[√]** where appropriate.
(1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

No.	Statement	1	2	3	4	5
D1	My organization regularly provides training on phishing awareness and cybersecurity practices.	□	□	□	□	□
D2	Senior management shows a strong commitment to cybersecurity priorities and risk mitigation.	□	□	□	□	□
D3	Employees in our institution frequently click suspicious or unknown email links.	□	□	□	□	□
D4	Internal audits or risk assessments effectively address cybersecurity concerns.	□	□	□	□	□
D5	Employees follow password best practices (e.g., strong, unique passwords).	□	□	□	□	□

SECTION E: AI READINESS AND ADOPTION

22. **Are you familiar with the concept of using Artificial Intelligence (AI) for cybersecurity purposes (e.g., phishing detection, behavioral analysis)?**

☐ Yes, I am familiar with such systems

☐ I have heard of it, but do not understand it well

☐ No, I am not familiar with AI for cybersecurity

23. **Would your organization consider adopting an AI-based system to detect phishing attacks?**

☐ Yes, we are open to such technologies

☐ No, due to cost or complexity

☐ Maybe, depending on further understanding and proof of effectiveness

24. **What concerns, if any, would you have about implementing AI in your cybersecurity operations?** (Check all that apply)

☐ Cost of implementation and maintenance

☐ Lack of technical expertise internally

☐ Trust in AI decisions and alert accuracy

☐ Integration challenges with legacy systems

☐ Data privacy and compliance risks

☐ Lack of transparency or explainability

25. **Would your institution be willing to participate in a pilot project involving an AI-driven phishing detection system?**

☐ Yes, we would be willing to participate

☐ No, not at this time

☐ Maybe, depending on the scope and resources required

SECTION F: PERCEPTIONS ON READINESS TO ADOPT AI-BASED PHISHING MITIGATION TOOLS

26. Please indicate your level of agreement with the following statements related to the adoption of AI-based phishing detection tools in your institution. Tick [√] where appropriate

(1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree)

No.	Statement	1	2	3	4	5
F1	I believe AI-based phishing detection would improve our cybersecurity operations.	☐	☐	☐	☐	☐
F2	I feel confident that our institution has the capacity to adopt AI-based systems.	☐	☐	☐	☐	☐
F3	I trust AI-generated phishing alerts as reliable.	☐	☐	☐	☐	☐

F4	AI systems may create new privacy or compliance challenges for MFIs.	☐	☐	☐	☐	☐
F5	Our staff would be open to using AI-based alerts or tools.	☐	☐	☐	☐	☐

SECTION G: GUIDED OPEN-ENDED RESPONSES

27. In your opinion, what do you think would be the most pressing cybersecurity threats facing MFIs today? (You may choose from the list below and add others)

☐ Phishing and social engineering attacks

☐ Ransomware

☐ Insider threats

☐ Credential theft and account compromise

☐ Malware and trojans

☐ Mobile money fraud

☐ Data breaches and unauthorized access

Other (please specify): _____

28. What specific steps do you think MFIs would take to better protect against phishing and social engineering attacks?

☐ Conduct regular cybersecurity training for staff

☐ Implement AI-based email and URL filtering

☐ Strengthen user access controls (e.g., 2FA, RBAC)

☐ Update cybersecurity policies and procedures

☐ Simulate phishing tests and response drills

☐ Improve endpoint security (antivirus, mobile security)

Other recommendations: _____

29. What policy or regulatory changes would you recommend to strengthen cybersecurity in the microfinance sector?

☐ Mandate cybersecurity compliance audits for MFIs

☐ Establish minimum cybersecurity certification for IT staff

☐ Enforce digital transaction monitoring and fraud alerts

☐ Provide subsidized access to cybersecurity tools or infrastructure

☐ Create an MFI-specific cybersecurity reporting framework

Other suggestions: _____

30. Which types of staff activities or departments in your institution do you think are the most frequently targeted by phishing attackers?

☐ Customer support / front desk

☐ Finance and loan processing

☐ IT or systems administration

☐ HR or recruitment

☐ Branch managers

Other (please specify): _____

31. Please share any additional comments or observations regarding phishing or cybersecurity in your institution

<< Thank you for your time and cooperation. Your resourceful contribution to this study is highly appreciated >>

Appendix 2: Interview Schedule

Interview Guide for MFIs on AI Adoption and Cybersecurity

Dear Participant,

I appreciate your participation in this interview. It is part of academic research on cybersecurity issues, phishing threats, and AI usage in microfinance institutions (MFIs). Your responses will remain private and confidential and will only be used for research purposes. Participation in this study is voluntary and anonymous, and your name will not be used in any part of the report.

Completing this interview should take about ten minutes.

Interviewer:	
Note-taker (if applicable):	
Date:	
Time Begin:	

Time End	
Location:	
Date of interview:	
Informed Consent (Sign if consent taken for all participants)	

Interview Questions

1. What are the primary challenges faced in managing cybersecurity incidents, and what are the current measures to address phishing attacks? (To check on Cybersecurity Challenges)

- *Follow-up Question*: Are there areas where these measures fall short?

2. Describe a recent phishing attack and the organizational response to the attack (To check on phishing incident analysis)

- *Follow-up Question*: Were there any changes made to policies or training after the incident?

3. What is the **r**ole of employees in phishing vulnerabilities? Is employee training effective? (To check on Employee Vulnerability)

- *Follow-up Question*: What improvements do you recommend for these programs?

4. Is AI adoption under consideration in your institution? (To check on AI Readiness)

- *Follow-up Question*: What specific AI tools have been explored, and what challenges have you faced in implementation?

5. Process for making cybersecurity investments. (To check on Organizational Policies)

- *Follow-up Question*: Are financial or technical constraints impacting these decisions?

6. What measures have been put in place to strengthen phishing prevention and detection (To check on Future Directions)

- *Follow-up Question*: What role should AI and regulatory bodies play in improving cybersecurity resilience?

7. Do you have any additional comments or recommendations?

End of Interview. Thank you for your time

Appendix 3: Public Datasets for GRU-Based Phishing Detection

Dataset Name	Source	Size	Key Features	Access Link
Phishing URL Dataset (PhishTank + Others)	Mendeley Data (2022)	88,647 URLs	Domain info, length, HTTPS, redirection flags, subdomain structure	Link
Enron-Spam & Phishing Email Corpus	Carnegie Mellon University	35,000+ labeled emails	Sender domain, subject line, attachments, body content, social engineering cues	Link
CIC-IDS2017 (Network Traffic Dataset)	Canadian Institute for Cybersecurity	2.8 million records	Flow duration, protocol flags, packet stats, timestamps, traffic labels	Link

Appendix 4: List of Licensed and authorized Saccos Societies in Nairobi, Kenya

SCHEDULE I: LIST OF SACCO SOCIETIES LICENSED TO UNDERTAKE DEPOSIT TAKING BUSINESS IN KENYA FOR THE PERIOD 1ST JANUARY 2024 TO 31ST DECEMBER 2024						
SNO	**Names of the Deposit Taking SACCO Society**	**Description**	**Key Products Offered**	**Postal Address**	**Physical Location of Head Office**	**County Location of Head Office**
1	Acumen Sacco Society Limited	A fast-growing SACCO serving individuals and groups in business, corporate, and the general public, with a focus on financial growth and investment. It functions as a key financial intermediary within a highly competitive and dynamic financial landscape.	FOSA Services, BOSA Loans, Business Loans, Investment Accounts, Mobile Banking	P.O. Box 1325 – 00200, Kajiado	Great Wide Mall, Magadi Road, Nairobi	Nairobi
2	Afya Sacco Society Ltd	Primarily serves professionals in the healthcare sector, including doctors, nurses, clinical officers, and medical support staff across Kenya. It represents a professional association leveraging the cooperative model for the	FOSA Services, Medical Equipment Loans, Development Loans, School Fees	P.O Box 11607 – 00400, Nairobi.	Afya Centre, Tom Mboya Street, Nairobi	Nairobi

		economic benefit of its members in the essential health services industry.	Loans, Salary Advances			
3	Airports Sacco Society Ltd	Caters to employees of the Kenya Airports Authority (KAA) and other stakeholders within the aviation and air transport industry. The SACCO supports the financial well-being of personnel critical to national and international air transport connectivity.	FOSA Services, BOSA Loans, Asset Financing, Emergency Loans, Holiday Packages	P.O Box 19048 – 00500, Nairobi	Kenya Airports Authority Head Quarters Building	Nairobi
4	Ardhi Sacco Society Ltd	Serves professionals in the land and geospatial sectors, including surveyors, cartographers, physical planners, and staff from the Ministry of Lands. It facilitates capital accumulation and credit access for professionals engaged in national land administration and spatial planning.	FOSA Services, Land Purchase Loans, Survey Equipment Financing, Development Loans	P.O Box 28782 – 00200, Nairobi	Survey Kenya Field Headquarters, Off Thika Road	Nairobi
5	Asili Sacco Society Ltd	A well-established SACCO with a diverse membership base, originally founded by a group of cooperative pioneers. Its long-standing operation exemplifies the resilience and enduring relevance of the cooperative movement in Kenya's financial sector.	FOSA Services, BOSA Loans, Savings Accounts, Mortgage Loans	P.O Box 49064 – 00100 Nairobi	Asili Co – op Centre, The Lower Ngara Road, Nairobi	Nairobi

6	Chai Sacco Society Ltd	Originally founded for employees of the Kenya Tea Development Authority (KTDA), it now serves a wide range of members from various sectors. The SACCO maintains a strong historical link to the tea industry, a cornerstone of the Kenyan agricultural economy.	FOSA Services, Agri-Business Loans, BOSA Loans, Asset Finance, Mobile Banking	P.O Box 278 – 00200, Nairobi	Chai House, Koinange Street, Nairobi	Nairobi
7	Chuna Sacco Society Ltd	The official SACCO for the staff and students of the University of Nairobi, one of the largest university-based SACCOs in Kenya. It promotes a culture of financial prudence and savings within a premier institution of higher learning and research.	FOSA Services, Student Fee Loans, Research Grants, Staff Development Loans, Laptops & Books Loans	P.O Box 30197 – 00100, Nairobi	University Of Nairobi, Engineering Department Building – Harry Thuku Road	Nairobi
8	Defense Sacco Society Ltd	Exclusively for serving and retired members of the Kenya Defense Forces (KDF), Ministry of Defense civilians, and their families. This SACCO operates as a closed financial system providing tailored services to the nation's military personnel and their families.	FOSA Services, Home Ownership Schemes, Vehicle & Asset Finance, Insurance Services (Bancassurance)	P.O Box 40668 – 00100 Nairobi	Ulinzi House, Lenana Road, Hurlingham, Nairobi	Nairobi
9	Elimu Sacco Society Ltd	Caters to individuals and groups within the education sector, including teachers, lecturers, and support staff in educational institutions. It is	FOSA Services, School Fees Loans, Teacher	P.O. Box 10073 –	Mubiru Road, Nairobi	Nairobi

		dedicated to empowering those responsible for the nation's educational development and knowledge dissemination.	Professional Development Loans, Construction Loans	00100, Nairobi		
10	Harambee Sacco Society Ltd	One of Kenya's largest and oldest SACCOs, originally for civil servants but now open to all, with a strong national presence. Its extensive branch network and large asset base position it as a systemically important player in the Kenyan cooperative sector.	FOSA Services, Jijenge Loans (Development), Mortgage Finance, Asset Financing, SaccoLink ATM	P.O Box 47815 – 00100, Nairobi	Harambee Sacco Plaza, Haile Selassie Avenue, Nairobi	Nairobi
11	Hazina Sacco Society Ltd	Open to the general public with a focus on wealth creation, it has a significant membership from both the corporate and business sectors. The SACCO's name and mission are directly aligned with the core objective of asset and wealth accumulation for its members.	FOSA Services, Wealth Creation Loans, Investment Accounts, Business Loans, Property Finance	P.O Box 59877 – 00200, Nairobi	Hazina Sacco Place, Ngong Road, Nairobi	Nairobi
12	Jamii Sacco Society Ltd	A community-focused SACCO that serves a diverse membership from various professions and small business owners. Its name, translating to "Society," reflects its foundational principle	FOSA Services, Jamii Loan (Development), Emergency Loans,	P.O. Box 57929 – 00200, Nairobi	Jamii Sacco Court, Mukenia Road, South B, Nairobi	Nairobi

		of serving the broader community through financial inclusion.	Micro-Enterprise Loans (Vikoba)			
13	Jogoo Sacco Society Ltd	Serves a broad membership base, with a significant number of members from the public service and corporate entities. It operates as a key financial intermediary, mobilizing savings and providing credit across a wide spectrum of the urban economy.	FOSA Services, Development Loans, School Fees Financing, Emergency Advances, SaccoLink	P.O Box 56074 - 00200, Nairobi	Commodore Office Suites, Off Ngong/Ring Road Kindaruma Avenue, Nairobi	Nairobi
14	Kencream Sacco Society Ltd	Originally established for employees of the Kenya Co-operative Creameries (KCC), it now also serves the wider community. The SACCO has historical roots in the dairy industry, a vital component of the Kenyan agricultural sector.	FOSA Services, Dairy Farming Loans, BOSA Loans, Savings and Credit Products	P.O Box 30131 – 00100, Nairobi	Creamary House, Dakar Road, Nairobi	Nairobi
15	Kenpipe Sacco Society Ltd	Primarily serves the employees of the Kenya Pipeline Company, providing financial services to those in the energy sector. It supports the financial stability of personnel responsible for the critical national petroleum pipeline infrastructure.	FOSA Services, Development Loans, Emergency Loans, Asset Financing	P.O Box 314 – 00507, Nairobi	Kenpipe Plaza, Sekoni Road, Nairobi	Nairobi

16	Kenversity Sacco Society Ltd	A SACCO dedicated to the university community, including staff and students from various public and private universities in Kenya. It fosters financial literacy and access to credit within the academic ecosystem nationwide.	FOSA Services, Student Accommodation Loans, Research & Conference Loans, Laptop Financing	P.O Box 10263 – 00100, Nairobi	Kenversity Plaza, Kahawa Sukari Road, Nairobi	Nairobi
17	Kenya National Police DT Sacco Society Ltd	The premier SACCO for serving and retired officers of the National Police Service, with a massive membership across the country. It represents one of the largest uniformed services SACCOs, crucial for the welfare of law enforcement personnel.	FOSA Services, Home Ownership Plan (HOP), Car Loan Scheme, SaccoLink, Insurance	P.O Box 51042 – 00200, Nairobi	Kenya Police Sacco Plaza, Ngara Road, Nairobi	Nairobi
18	Kimisitu Sacco Society Ltd	Caters primarily to employees of non-governmental organizations (NGOs), diplomatic missions, and international development agencies. It serves a unique membership base engaged in the national and international development sector.	FOSA Services, Development Loans, Foreign Currency Accounts, School Fees Loans, Travel Loans	P.O. Box 10454 – 00200, Nairobi	AEA Plaza, Valley Road, Nairobi	Nairobi
19	Magereza Sacco Society Ltd	Serves the uniformed staff of the Kenya Prisons Service, both serving and retired, and their families. This SACCO provides a dedicated	FOSA Services, Development Loans, Emergency	P.O. Box 53131 –	Mageso Chambers, Moi	Nairobi

		financial services channel for personnel within the corrections and rehabilitation service.	Loans, Asset Finance, Welfare Products	00200, Nairobi	Avenue, Nairobi	
20	Maisha Bora Sacco Society Ltd	Open to all, this SACCO focuses on empowering its members from all walks of life to achieve a "good life" through financial services. Its name and mission are directly focused on the socio-economic upliftment of its membership.	FOSA Services, Maisha Bora Loan, Business Loans, School Fees Financing	P.O. Box 30062 – 00200, Nairobi	Commercial Street, Industrial Area, Nairobi	Nairobi
21	Mwalimu National Sacco Society Ltd.	The largest teachers' SACCO in Kenya, exclusively serving teachers in public and private institutions, with immense financial strength. It is a systemically important institution within the cooperative sector, wielding significant economic influence.	FOSA Services, Mwalimu Mortgage, Car Loan Scheme, Mwalimu SaccoLink, Biashara Loans	P.O. Box 62641 – 00200, Nairobi	Hill Lane, Nairobi City	Nairobi
22	Mwito Sacco Society Ltd	A SACCO with a diverse membership base, known for serving small and medium-sized enterprise (SME) owners and individuals. It plays a significant role in providing financial solutions to the entrepreneurial sector in Nairobi.	FOSA Services, SME Loans, Development Loans, Quick Emergency Loans	P.O Box 56763 – 00200, Nairobi	Mwito Building, Desai Road, Nairobi	Nairobi

23	Nacico Sacco Society Ltd	Serves the employees of the National Council of Churches of Kenya (NCCK) and its member churches and affiliated organizations. It operationalizes the church's social ministry through structured financial intermediation for its staff and affiliates.	FOSA Services, Church Project Financing, Development Loans, Clergy Support Loans	P.O Box 34525 – 00100, Nairobi	NACICO Plaza, Landhies Road, Nairobi	Nairobi
24	Nafasi Sacco Society Ltd	Caters to the employees of the National Cereals and Produce Board (NCPB) and other members from the agricultural sector. The SACCO supports the financial needs of personnel central to national food security management.	FOSA Services, Agri-Loans, Grain Storage Financing, Development Loans	P.O. Box 41426 – 00100 Nairobi	NCPB Nairobi Silos Complex, Off Outering Road, Nairobi	Nairobi
25	Nation DT Sacco Society Ltd	Originally for employees of the Nation Media Group, it has expanded to serve a wide range of members from the public and private sectors. It has evolved from a staff association to a broad-based financial services provider.	FOSA Services, Development Loans, Instant Salary Advances, Business Loans	P.O. Box 22022 – 00400, Nairobi	Cambrian Building, Moi Avenue, Nairobi	Nairobi
26	Njiwa Sacco Sacco Ltd	A long-standing SACCO with a diverse membership, popular among small business owners and salaried employees in Nairobi. It exemplifies the classic community-based model that aggregates savings and provides credit within an urban setting.	FOSA Services, Njiwa Loan, Business Quick Loans, Emergency Savings	P.O Box 10221 - 00100, Nairobi	Golf View Office Suites, Wambui Road, Nairobi	Nairobi

27	NSSF Sacco Society Ltd	Serves the employees of the National Social Security Fund (NSSF), providing tailored financial products for its members. The SACCO supports the staff of the national institution mandated with social security protection.	FOSA Services, Development Loans, Top-Up Loans, Retirement Planning Products	P.O Box 43338 – 00100, Nairobi	Social Security House, Bishops Road, Nairobi	Nairobi
28	Nyati Sacco Society Ltd	A SACCO that serves a mixed membership from various professions and businesses, promoting a culture of saving and investment. Its name, symbolizing strength, reflects its mission to be a robust financial partner for its members.	FOSA Services, Nyati Loan (Development), Asset Finance, School Fees Loans	P.O Box 7601 – 00200, Nairobi	Odyssey Plaza, Mukoma Road, Nairobi	Nairobi
29	Safaricom Sacco Society Ltd	Originally for Safaricom employees, it has grown into a major SACCO also serving the general public with innovative, tech-driven services. It leverages technological innovation to enhance member service delivery and operational efficiency.	FOSA Services, M-Pesa Integrated Loans, M-Shwari-like Products, School Fees Loans, Asset Finance	P.O Box 6682 – 00800, Nairobi	Safaricom Care Centre, Waiyaki Way, Westlands, Nairobi	Nairobi
30	Sheria Sacco Society Ltd	The SACCO for professionals in the legal field, including lawyers, magistrates, judges, and law office support staff. It constitutes a professional financial collective for practitioners and officers of the court.	FOSA Services, Law Firm Financing, Case Cost Loans, Professional	P.O. Box 34390 – 00100, Nairobi	Matumbato Close, Nairobi	Nairobi

			Development Loans			
31	Shirika DT Sacco Society Ltd	Serves employees of state corporations and parastatals ("Shirika") across various government ministries and agencies. It is a key financial vehicle for staff within the public corporate sector in Kenya.	FOSA Services, Development Loans, Emergency Loans, Group Savings Schemes	P.O Box 43429 – 00100, Nairobi	Shirika Coop House, Ngara Kipande Road, Ngara, Nairobi	Nairobi
32	Shoppers Sacco Society Ltd	A SACCO with a diverse membership, known for its services to traders, merchants, and salaried employees in the urban center. It caters specifically to the financial dynamics of the retail and commerce sector.	FOSA Services, Trade Stock Loans, Quick Business Loans, Instant Savings	P.O BOX 16-00507, Nairobi	Nature House – Tom Mboya Street	Nairobi
33	Stima DT Sacco Society Ltd	One of the leading SACCOs in Kenya, originally for Kenya Power employees but now a major player serving all sectors of the economy. It has successfully transitioned from an occupational SACCO to a dominant national financial intermediary.	FOSA Services, Stima Loan, Imarisha Loan (Asset Finance), Stima Sacco Mobile Banking	P.O Box 75629 – 00200, Nairobi	Stima Sacco Plaza, Mushembi Road, Parklands, Nairobi	Nairobi
34	Strategic DT Sacco Society Ltd.	Focuses on providing strategic financial solutions to a diverse membership, including corporates, SMEs, and individuals. Its name reflects a deliberate approach to crafting	FOSA Services, Strategic Investment Loans, Customized	P.O. Box 78506 – 00507, Nairobi	Lunga Lunga Square, Industrial Area,	Nairobi

		financial products that meet specific member goals.	Business Finance, Development Loans		Nairobi County.	
35	Taqwa Sacco Society Ltd	An Islamic-based SACCO that operates on Sharia-compliant principles, serving both Muslims and non-Muslims. It fills a unique niche in the market by offering interest-free financial products that comply with Islamic finance principles.	Sharia-Compliant Financing (Murabaha, Ijara), Sharia-Compliant Savings Products	P.O Box 10180 – 00100, Nairobi	Jamia Plaza, Kigali Street, Nairobi	Nairobi
36	Telepost Sacco Society Ltd	Serves employees of Telkom Kenya and the Postal Corporation of Kenya, as well as their immediate family members. This SACCO has its roots in the historic telecommunications and postal services sector of Kenya.	FOSA Services, Development Loans, Emergency Advances, Staff Welfare Loans	P.O Box 49557 – 00100, Nairobi	City Square Post Office, Haile Selassie Avenue, Nairobi	Nairobi
37	Tembo Sacco Society Ltd	A growing SACCO serving individuals and groups, with a strong focus on personal development and business financing. Its name, symbolizing strength and memory, reflects its commitment to being a steadfast financial partner.	FOSA Services, Tembo Loan (Development), Business Loans, School Fees Financing	P.O Box 91 – 00618, Ruaraka	Tembo Sacco Complex, Garden Estate Road, Mukima Drive, Nairobi	Nairobi

38	The Kenya Bankers Sacco Society Ltd.	The premier SACCO for professionals working in the banking industry in Kenya, from tellers to senior management. It represents a financial cooperative for professionals who are themselves experts in finance and banking.	FOSA Services, Bankers' Mortgage, Car Loans, Personal Loans, High-Yield Investment Accounts	P.O. Box 73236 – 00200, Nairobi	The Kenya Bankers Sacco Centre, 3rd Ngong Ave., Nairobi	Nairobi
39	Ufanisi DT Sacco Society Ltd	A SACCO whose name means "Development," focusing on empowering members through loans for business, home ownership, and education. Its entire operational philosophy is oriented towards fostering member and community development.	FOSA Services, Ufanisi Loan (Development), Business Loans, Plot & House Construction Loans	P.O. Box 2973 – 00200, Nairobi	Development House, Moi Avenue, Nairobi	Nairobi
40	Ukristo Na Ufanisi Sacco Society Ltd	A faith-based (Christian) SACCO that integrates faith with financial empowerment for its members, often linked to local churches. It represents a model of holistic ministry addressing both spiritual and economic needs.	FOSA Services, Church Project Loans, Development Loans, Welfare & Benevolence Funds	P.O Box 872 – 00605 Nairobi	Ack Emmanuel Church Riruta, Nairobi	Nairobi

41	Ukulima Sacco Society Ltd	One of Kenya's oldest SACCOs, originally for farmers and agricultural cooperatives, now serving a wide array of members. It maintains a strong legacy and connection to the agricultural sector, the backbone of the Kenyan economy.	FOSA Services, Agri-Loans, Development Loans, LPO Financing, SaccoLink	P.O Box 44071 – 00100, Nairobi	Ukulima Cooperative House, Haile Selassie Avenue, Nairobi	Nairobi
42	Unaitas Sacco Society Ltd	A leading national SACCO with a strong focus on financial inclusion, serving members in both urban and rural areas with a wide branch network. Its innovative approach has made it a significant player in deepening financial access in Kenya.	FOSA Services, Unaitas Biashara Loan, Unaitas Stude Loan, Unaitas Poultry Loan, Mobile Banking	P.O Box 38721 – 00100, Nairobi	Cardinal Otunga Plaza, Nairobi	Nairobi
43	United Nations DT Sacco Society Ltd	Exclusively for the staff of the United Nations and its affiliated agencies based in Kenya. It is a unique, internationally-focused SACCO serving a diverse and multicultural membership from the diplomatic community.	FOSA Services, Foreign Currency Denominated Loans, International School Fees Financing, Travel Loans	P.O Box 2210 – 00621, Nairobi	United Nations Complex, United Nations Avenue, Nairobi	Nairobi

44	Ushuru Sacco Society Ltd	Serves the employees of the Kenya Revenue Authority (KRA), offering financial products tailored to civil servants. The SACCO supports the financial needs of personnel tasked with mobilizing public revenue for the government.	FOSA Services, Development Loans, Ushuru Jipange Loan (Asset Finance), Mortgage Top-Up	P.O Box 52072 – 00200, Nairobi	Ushuru Sacco Centre, Wood Avenue, Kilimani	Nairobi
45	Wanaanga Sacco Society Ltd	Caters to the staff of the Kenya Meteorological Department and other professionals in the environmental and climate sciences. It serves experts in the critical field of weather, climate, and atmospheric sciences.	FOSA Services, Development Loans, Emergency Loans, Weather Station Equipment Financing	P.O Box 34680 – 00100, Nairobi	Kenya Meteorological Department,NgongRoad,Dagoretti	Nairobi
46	Wanandege Sacco Society Ltd	Serves a wide range of professionals in the aviation industry, including pilots, engineers, cabin crew, and airport operations staff. The SACCO is the primary financial cooperative for Kenya's aviation professionals.	FOSA Services, Aircraft Maintenance Loans, Pilot Training Financing, Travel & Per Diem Advances	P.O Box 19074 – 00501, Nairobi	Wanandege Plaza, Old North Airport Road, Nairobi	Nairobi
47	Waumini Sacco Society Ltd	The official SACCO of the Catholic Church in Kenya, serving clergy, religious institutions,	FOSA Services, Church	P.O Box 66121 –	Applewood Adams 2nd	Nairobi

		and the lay faithful across the country. It is one of the largest faith-based SACCOs, integrating Catholic social teaching with financial services.	Construction Loans, Clergy Support Loans, Development Loans, Waumini Mobile	00800, Nairobi	Floor, Ngong Road, Nairobi	

SCHEDULE II: LIST OF SACCO SOCIETIES AUTHORIZED TO UNDERTAKE SPECIFIED NON-DEPOSIT-TAKING BUSINESS (BOSA ONLY) FOR THE PERIOD 1ST JANUARY 2024 TO 31ST DECEMBER 2024						
SNO	**Names of the Regulated Non-With-Drawable Deposit Taking Sacco Society**	**Description**	**Key Products Offered**	**Postal Address**	**Physical Location of Head Office**	**County Location of Head Office**
1	Communication Regulated Non-WDT-Sacco	Serves employees of the Communication Authority of Kenya and the broader telecommunications sector. It functions as a key financial intermediary within a highly regulated and technologically dynamic industry.	BOSA Loans, Development Loans, School Fees Loans, Emergency Loans	P.O. Box 1677 - 00606, Nairobi	Communication Authority Centre, Westlands	Nairobi
2	Concorde Regulated Non-WDT-Sacco	A SACCO serving a diverse membership, potentially linked to a specific organization or community. Its operational model exemplifies the application of cooperative principles within a targeted demographic.	BOSA Loans, Savings Products, Asset Financing	P.O.Box 10690 - 00100, Nairobi	Centenary House, Westlands	Nairobi
3	Co-operative Bank Regulated Non-WDT-Sacco	Primarily for the staff of Co-operative Bank of Kenya and their families. This structure represents a financial services entity operating a	BOSA Loans, Emergency Loans, School Fees	P.O. Box 48231 - 00100, Nairobi	Co-Op Bank House, Nairobi CBD	Nairobi

		parallel internal thrift and credit cooperative for its workforce.	Financing, Salary Advances			
4	Davis & Shirtliff Regulated Non-WDT-Sacco	Serves the employees of Davis & Shirtliff, a leading supplier of water and energy equipment. The SACCO contributes to employee financial stability within the essential water technology and infrastructure sector.	BOSA Loans, Development Loans, Emergency Credit, Group Savings	P.O Box 41762 - 00100, Nairobi	Davis & Shirtliff Building, Dundori Road, Industrial Area	Nairobi
5	Devco Regulated Non-WDT-Sacco	Likely serves employees in the development and construction sectors. It facilitates capital accumulation and credit access for professionals engaged in national physical infrastructure development.	BOSA Loans, Project Financing, Land Purchase Loans	P.O Box 30645 - 00100, Nairobi	KMA Building 7th Floor, Mara Road Upper Hill	Nairobi
6	Dhamini Regulated Non-WDT-Sacco	A community-based SACCO; "Dhamini" translates to "contributor" or "depositor". Its nomenclature directly reflects the fundamental cooperative principle of member savings and collective resource pooling.	BOSA Loans, Daily Savings, Emergency Loans, Micro-Loans	P.O Box 44749 - 00100, Nairobi	ACK Parking Silo, Bishop Road, Westlands	Nairobi
7	DHL Regulated Non-WDT-Sacco	For employees of DHL, the global logistics and courier company, in Kenya. This SACCO is integral to providing financial services within a	BOSA Loans, Logistical Asset Finance,	P.O Box 67577 -	DHL House, Witu Road Off Lusaka	Nairobi

		multinational corporate environment's local operational framework.	Emergency Advances	00200, Nairobi	Road, Industrial Area	
8	Digital Media Regulated Non-WDT-Sacco	Caters to professionals and businesses in the digital media and creative industry. It supports a rapidly evolving economic sector characterized by project-based work and entrepreneurial ventures.	BOSA Loans, Business Loans for SMEs, Equipment Financing	P.O Box 65825 - 00400, Nairobi	2nd Floor, Red Commercial Complex, Embakasi South	Nairobi
9	Dudu Regulated Non-WDT-Sacco	Likely serves employees in the agricultural or environmental sector ("Dudu" means insect). Its operational scope is closely aligned with entomological research and agricultural pest management initiatives.	BOSA Loans, Agricultural Input Loans, Savings Products	P.O Box 30772 - 00100, Nairobi	TRO ConferenceBuilding,ICIPE,Kasarani	Nairobi
10	Embassava Regulated Non-WDT-Sacco	A large transport SACCO originally linked to the matatu industry on the Embakasi-Avalon route. It represents a significant financial mobilization vehicle within the informal public transport ecosystem.	BOSA Loans, Matatu Financing, Insurance, Daily Savings	P.O Box 3546 - 00200, Nairobi	Jiwabhai Vekaria Building, Taveta Road, Nairobi CBD	Nairobi
11	Energy Regulated Non-WDT-Sacco	Serves professionals and employees in the energy and petroleum sectors. The society plays a crucial role in the financial inclusion of	BOSA Loans, Energy Project	P.O. Box 10585 -	Kawi Complex	Nairobi

		personnel powering the nation's critical energy infrastructure.	Loans, Asset Finance	00100, Nairobi	Block A, South C	
12	Equity Regulated Non-WDT-Sacco	For employees of Equity Group Holdings, including Equity Bank. It operates as an internal financial cooperative for the staff of a leading financial institution renowned for its social banking model.	BOSA Loans, Staff Advances, Mortgage Top-up Loans	P.O. Box 75104 - 00200, Nairobi	Equity Centre, Upper Hill	Nairobi
13	Exams Regulated Non-WDT-Sacco	Serves employees of the Kenya National Examinations Council (KNEC). This SACCO supports the financial well-being of professionals tasked with upholding the integrity of the national education assessment system.	BOSA Loans, School Fees Loans, Development Loans	P.O Box 73598 - 00200, Nairobi	NewMitihani House,Popo Road, South C	Nairobi
14	Family Regulated Non-WDT-Sacco	For employees of Family Bank and their immediate family members. Its structure underscores the prevalence of staff-focused financial cooperatives within the Kenyan banking industry.	BOSA Loans, Instant Salary Advances, Family Emergency Loans	P.O Box 38993 - 00100, Nairobi	10th Family Bank Towers, Nairobi CBD	Nairobi
15	Faulu Bank Regulated Non-WDT-Sacco	Serves employees of Faulu Microfinance Bank. The SACCO exemplifies the practice of microfinance institutions establishing their own internal member-based financial services associations.	BOSA Loans, Micro-Savings, Emergency Credit	P.O Box 60240 - 00200, Nairobi	Faulu Microfinance Bank Building, Off	Nairobi

					Ngong Road, Nairobi	
16	Finnlemm Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. It operates as a classic example of a common-bond SACCO that brings together individuals from various professional and social backgrounds.	BOSA Loans, Investment Loans, Land Buying Loans	P.O Box 67666 - 00200, Nairobi	Finnlemm House, 55 Gatundu Road, Kileleshwa, Nairobi	Nairobi
17	Fugo Regulated Non-WDT-Sacco	A faith-based SACCO associated with the Full Gospel Churches of Kenya. It integrates theological principles with economic empowerment, fostering a community of shared belief and financial mutualism.	BOSA Loans, Church Project Loans, Welfare Loans	P.O Box 6434 - 00300, Nairobi	Full Gospel Churches of Kenya, Along Manyanja Road	Nairobi
18	Grand Granite Diaspora Regulated Non-WDT-Sacco	Caters specifically to Kenyans in the diaspora, particularly those with connections to the "Grand Granite" group. It leverages transnational networks to channel remittances into productive investment and savings.	Diaspora Remittance Loans, Property Investment Loans, BOSA Savings	P.O Box 18399 - 00100, Nairobi	KMA APT, Block A, Suite 2.2, Chyulu Road, Upper Hill	Nairobi
19	Haco Regulated Non-WDT-Sacco	Serves employees of Haco Industries, a major manufacturer of consumer brands. The SACCO functions as a key employee benefit scheme	BOSA Loans, Production Input Financing, Emergency Loans	P.O Box 43903 - 00100, Nairobi	Kasarani Mwiki Road,	Nairobi

		within the fast-moving consumer goods (FMCG) manufacturing sector.				
20	Heart Regulated Non-WDT-Sacco	A community-focused SACCO, potentially with a membership base in the Karen area. It exemplifies a residential or neighborhood-based common bond for fostering local financial resilience.	BOSA Loans, Home Improvement Loans, School Fees Financing	P.O Box 70000-00400, Nairobi	Tana House, Karen Shopping Centre, Nairobi	Nairobi
21	HELB Regulated Non-WDT-Sacco	For employees of the Higher Education Loans Board. This SACCO serves the institution responsible for financing human capital development, thereby supporting its own staff's financial development.	BOSA Loans, Education Loans, Asset Financing	P.O Box 69489 - 00400, Nairobi	Anniversary Towers, Nairobi CBD	Nairobi
22	Hoechem Regulated Non-WDT-Sacco	Serves employees of Hoechem, a company in the chemical industry. It provides tailored financial services to professionals within a specialized and capital-intensive industrial segment.	BOSA Loans, Chemical Input Financing, Equipment Loans	P.O Box 30467 - 00100, Nairobi	Highchem Complex, Mogadishu Road, Industrial Area, Nairobi	Nairobi
23	ICEA Agents Regulated Non-WDT-Sacco	For insurance agents and staff associated with the ICEA Lion Group. The SACCO is designed to cater to the unique income streams and	BOSA Loans, Commission Advance Loans, Business Loans	P.O Box 46143 - 00100, Nairobi	Nairobi West, Kodi Road House No. 24, Nairobi	Nairobi

		financial needs of commission-based insurance sales agents.				
24	Irrigation Regulated Non-WDT-Sacco	Serves employees of the National Irrigation Authority and related water sectors. Its membership is central to the implementation of national food security and agricultural water management policies.	BOSA Loans, Irrigation Equipment Loans, Agri-input Financing	P.O Box 30372 - 00100, Nairobi	Irrigation House, Lenana Road, Kilimani Area, Nairobi	Nairobi
25	Jachin Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. Its operations are indicative of the broad-based appeal and adaptability of the cooperative savings model across different societal segments.	BOSA Loans, Development Loans, Emergency Savings	P.O Box 12502 - 00400, Nairobi	Hadada Apartments, Kindaruma Road, Off Ngong Road	Nairobi
26	Java Regulated Non-WDT-Sacco	For employees of the Java House chain of restaurants and its affiliates. This SACCO supports workforce stability in the competitive hospitality and food service industry.	BOSA Loans, Hospitality Business Loans, Emergency Cash	P.O Box 21533 - 00505, Nairobi	ABC Place, WaiyakiWay, Westlands, Nairobi	Nairobi
27	Jumuia Ya Ulaya Regulated Non-WDT-Sacco	A diaspora-focused SACCO for Kenyans living in European countries ("Ulaya" means Europe). It represents a formal institutional mechanism for harnessing diaspora savings for development.	Diaspora Mortgage Loans, Investment Loans, BOSA Savings	P.O Box 29960 - 00100, Nairobi	Dunhill Towers, 20th Floor, Waiyaki Way,	Nairobi

					Westlands, Nairobi	
28	KAG Regulated Non-WDT-Sacco	Associated with the Kenya Assemblies of God church and its members. It embodies the fusion of faith-based community organization with structured financial self-help initiatives.	BOSA Loans, Church Development Loans, Welfare Loans	P.O Box 67014 - 00200, Nairobi	KAG East University, Buruburu, Nairobi	Nairobi
29	Kanisa Regulated Non-WDT-Sacco	A general church-based SACCO ("Kanisa" means Church), hosted at the AACC. Its location within an ecumenical body highlights the role of religious institutions in promoting socioeconomic empowerment.	BOSA Loans, Clergy Support Loans, Project Financing	P.O Box 1225 - 00606, Nairobi	AACC Compound, WaiyakiWay, Westlands	Nairobi
30	KASNEB Regulated Non-WDT-Sacco	For employees of the Kenya Accountants and Secretaries National Examinations Board. The SACCO serves professionals who are themselves custodians of accounting and secretarial standards.	BOSA Loans, Professional Development Loans	P.O Box 41360 - 00100, Nairobi	KASNEB Towers, Upper Hill, Nairobi	Nairobi
31	KEMRI Regulated Non-WDT-Sacco	Serves staff of the Kenya Medical Research Institute. It functions as a vital financial support system for individuals engaged in critical public health and scientific research.	BOSA Loans, Research Grant Advance Loans, Emergency Funds	P.O Box 19643 - 00202, Nairobi	Kemri Visiting Scientists Quarters,	Nairobi

					Mbagathi Road	
32	Kenchic Regulated Non-WDT-Sacco	For employees of Kenchic, a major poultry processing company. The SACCO is embedded within the agribusiness value chain, specifically the poultry and livestock sub-sector.	BOSA Loans, Poultry Farming Loans, Input Financing	P.O Box 20052 - 00200, Nairobi	Exsan House Building, Industrial Area	Nairobi
33	Kenred Regulated Non-WDT-Sacco	Serves employees and members of the Kenya Red Cross Society. It supports the financial welfare of personnel dedicated to humanitarian action and disaster response.	BOSA Loans, Emergency Response Loans, Welfare Credit	P.O Box 40712 - 00100, Nairobi	KenyaRedCrossBuilding, SouthC,Nairobi	Nairobi
34	Kentours Regulated Non-WDT-Sacco	For employees in the tourism and travel sector. The SACCO provides a financial buffer for workers in an industry known for its susceptibility to global economic and seasonal fluctuations.	BOSA Loans, Travel Loan Facilities, Vehicle Financing	P.O Box 79333 - 00200, Nairobi	Commodore Office Suites, Kindaruma Road, Kilimani,Nairbi	Nairobi
35	Kenya Medical Association Regulated Non-WDT-Sacco	For registered members of the Kenya Medical Association (doctors). It represents a professional association leveraging the cooperative model for the economic benefit of its members.	BOSA Loans, Medical Equipment Financing, Practice Loans	P.O Box 413 - 00202, Nairobi	KMA Centre, Upper Hill, Nairobi	Nairobi

36	Kenya Re Regulated Non-WDT-Sacco	For employees of the Kenya Reinsurance Corporation. The SACCO operates within the core of the financial services sector, serving experts in risk transfer and capital management.	BOSA Loans, Investment Loans, Asset Financing	P.O Box 30271 - 00100, Nairobi	Re-insurance Plaza, Nairobi CBD	Nairobi
37	Kenya Rural Roads Regulated Non-WDT-Sacco	Serves employees of the Kenya Rural Roads Authority (KeRRA). It is instrumental for the financial health of those responsible for developing and maintaining rural transport infrastructure.	BOSA Loans, Development Loans, Machinery Loans	P.O Box 48151 - 00100, Nairobi	Barabara Plaza, Airport SouthRoad, JKIA	Nairobi
38	Kenyatta Matibabu Regulated Non-WDT-Sacco	Associated with the staff of Kenyatta National Hospital ("Matibabu" means treatment). This SACCO is a cornerstone of financial support for frontline workers in the largest public referral hospital.	BOSA Loans, Medical Emergency Loans, Development Loans	P.O Box 2117 - 00202, Nairobi	Flamingo Towers, Along Mara Road, Upper Hill	Nairobi
39	Kewisco Regulated Non-WDT-Sacco	A SACCO with a diverse membership base, located in Nairobi West. It demonstrates the community-based model of SACCOs that serve a geographically defined urban population.	BOSA Loans, Business Loans, School Fees Loans	P.O Box 4491 - 00200, Nairobi	West End Place, Nairobi West	Nairobi
40	Kico Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. Its operational structure is representative of the many SACCOs that cultivate a broad-based	BOSA Loans, Personal Loans, Emergency Advances	P.O Box 55319 - 00200, Nairobi	Bruce House, Standard Street	Nairobi

		membership without a single occupational common bond.				
41	Kilele Regulated Non-WDT-Sacco	A SACCO operating from the KUSCCO Centre, serving a diverse membership. Its presence in the headquarters of the SACCO umbrella body signifies its integration into the broader cooperative movement.	BOSA Loans, Top-Up Loans, Investment Savings	P.O. Box 28403 - 00200, Nairobi	KUSCCO Centre, Along Kilimanjaro Avenue, Upper Hill,	Nairobi
42	Kinga Regulated Non-WDT-Sacco	A SACCO with a diverse membership; "Kinga" can mean "protect" or "shield". The name metaphorically reflects the SACCO's fundamental purpose of providing financial security and protection to its members.	BOSA Loans, Insurance-Linked Loans, Secure Savings	P.O. Box 22591 - 00400, Nairobi	Jumuia Place, Kilimani, Nairobi	Nairobi
43	Kingsize Regulated on-WDT-Sacco	Likely serves employees of Nairobi Bottlers (a Coca-Cola bottler). It functions as an internal financial vehicle within a major player in the beverage manufacturing and bottling industry.	BOSA Loans, Production Line Loans, Staff Advances	P.O Box 18034 - 00500, Nairobi	Nairobi Bottlers Premises, North Airport Road, Nairobi	Nairobi
44	Kirungii Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. It operates as a financial cooperative that pools	BOSA Loans, Jua Kali Business	P. O Box 30029 -	Deloitte Place, Waiyaki	Nairobi

		resources from individuals across various income-generating activities.	Loans, Market Loans	0100, Nairobi	Way, Muthangari Road, Nairobi	
45	Kumbukumbu Regulated Non-WDT-Sacco	A community-based SACCO; "Kumbukumbu" means "memory" or "memorial". The nomenclature may signify a foundational principle rooted in collective history or shared experience among its members.	BOSA Loans, Memorial Project Financing, Welfare Loans	P.O Box 22330 - 00400, Nairobi	Kipande Road, Museum Hill, Nairobi	Nairobi
46	Law Society Of Kenya Regulated Non-WDT-Sacco	For advocates who are members of the Law Society of Kenya (LSK). It constitutes a professional financial services arm for officers of the court and legal practitioners.	BOSA Loans, Practice Development Loans, Case Cost Financing	P.O Box 6740 - 00100, Nairobi	Crawford Business Park, Statehouse Road, Nairobi	Nairobi
47	Lompasago Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. Its structure is characteristic of urban SACCOs that aggregate savings and provide credit within a mixed-member common bond.	BOSA Loans, Micro-Business Loans, Chama Loans	P.O Box 17553 - 00500, Nairobi	Theevan Building, Jogoo Road	Nairobi
48	Madison Regulated Non-WDT-Sacco	For employees of Madison Group, an insurance and financial services company. The SACCO	BOSA Loans, Insurance Premium	P.O Box 47382 -	Madison House, Upper Hill	Nairobi

		represents a parallel financial structure for staff within a larger financial conglomerate.	Financing, Investment Loans	00100, Nairobi		
49	Maktaba Regulated Non-WDT-Sacco	Serves employees of the Kenya National Library Service ("Maktaba" means Library). It supports the financial inclusion of professionals dedicated to knowledge management and public literacy.	BOSA Loans, Book Acquisition Loans, Research Loans	P.O Box 4156 - 00200, Nairobi	Maktaba Kuu Building, Ngong Road	Nairobi
50	Masterways Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base, located in the CBD. Its central business district location facilitates access to financial services for a urban, commercially-oriented membership.	BOSA Loans, Commercial Paper Loans, Business Advances	P.O Box 38715 - 00600, Nairobi	Old Mutual Building, Kimathi Street, Nairobi CBD	Nairobi
51	Mhasibu Regulated Non-WDT-Sacco	A SACCO popular among accountants and finance professionals ("Mhasibu" means Accountant). It exemplifies a professional common bond centered on expertise in financial management and auditing.	BOSA Loans, Audit Fee Financing, Professional Development Loans	P.O Box 31295 - 00600, Nairobi	Absa Towers, Market Street, Nairobi	Nairobi
52	Minet Regulated Non-WDT-Sacco	For employees of Minet Kenya, an insurance brokerage and risk services firm. The SACCO provides a savings and credit platform for experts in corporate risk and insurance solutions.	BOSA Loans, Risk Cover Loans, Commission Advances	P.O Box 48279 - 00100, Nairobi	Minet House, Along Professional Way – Off Nyerere	Nairobi

					Road, Nairobi	
53	Mkombozi Regulated Non-WDT-Sacco	A SACCO serving a diverse membership; "Mkombozi" means "Saviour". The name connotes a foundational mission to provide financial liberation and rescue from economic precarity.	BOSA Loans, Debt Consolidation Loans, Emergency Rescue Loans	P.O Box 62727 - 00200, Nairobi	Alpha Centre, Godown No-85, Along Mombasa Road	Nairobi
54	Mofaa Regulated Non-WDT-Sacco	A community-based SACCO, likely in the Roysambu area. It serves as a localized institution for mobilizing domestic savings and stimulating economic activity within a specific residential zone.	BOSA Loans, Home Improvement Loans, Shop Financing	P.O Box 64742 - 00620, Nairobi	Jeda Plaza, Roysambu	Nairobi
55	Multiple Regulated Non-WDT-Sacco	Likely serves the transport and logistics sector, specifically multiple hauliers. It addresses the distinct capital and cash flow needs of players in the freight and goods movement industry.	BOSA Loans, Truck & Trailer Financing, Logistics Loans	P.O Box 41391 - 00100, Nairobi	Multiple Hauliers Building, AirportSouth Road	Nairobi
56	Mzima Springs Regulated Non-WDT-Sacco	For the staff and students of Strathmore University. This SACCO is integrated into the academic ecosystem, promoting a culture of financial prudence within a tertiary learning institution.	BOSA Loans, Student Fee Loans, Research Grants	P. O Box 59857 - 00200, Nairobi	Student Center Building, Strathmore University	Nairobi

57	Nairobi Consumers Regulated Non-WDT-Sacco	A SACCO serving traders and consumers in the Nairobi area, particularly around Gikomba. It is a key financial intermediary within one of East Africa's largest informal merchandise markets.	BOSA Loans, Trade Stock Loans, Mitumba Business Loans	P. O Box 36 - 00300, Nairobi	New Nafuu Classic Building, Popo Road, Gikomba	Nairobi
58	Nairobi Water Regulated Non-WDT-Sacco	For employees of the Nairobi Water and Sewerage Company. The SACCO supports the workforce responsible for the provision of essential water and sanitation services in the capital city.	BOSA Loans, Water Project Loans, Utility Asset Loans	P. O Box 13880 - 00100, Nairobi	Garden Chambers, Moktar Dadah Street, Nairobi	Nairobi
59	Network Regulated Non-WDT-Sacco	For employees of the Standard Group's media network. It functions as the internal financial cooperative for professionals in the news, broadcasting, and information dissemination sectors.	BOSA Loans, Media Equipment Loans, Travel Advance Loans	P.O Box 56985 - 00200, Nairobi	Standard Group Center, Mombasa Road	Nairobi
60	Nimepata Regulated Non-WDT-Sacco	A community-based SACCO; "Nimepata" translates to "I have received/gotten". The name embodies a tangible outcome-based philosophy, emphasizing the direct benefits of membership.	BOSA Loans, Instant Benefit Loans, Quick Savings	P.O Box 55405 - 00200, Nairobi	International Leadership University, Mtito Andei Road, Westlands	Nairobi

61	Nyumba - Nairobi Regulated Non-WDT-Sacco	A SACCO focused on housing and home ownership ("Nyumba" means House). Its strategic orientation is directly aligned with addressing the critical need for housing finance among its membership.	BOSA Loans, Home Construction Loans, Plot Purchase Loans	P.O Box 30088 - 00100, Nairobi	Applewood Adams, Ngong Road, Nairobi	Nairobi
62	Parents Plan Regulated Non-WDT-Sacco	Likely a SACCO focused on financial planning for families and education. It caters to the intergenerational financial goals of its members, particularly savings for children's education and welfare.	BOSA Loans, School Fees Planning Loans, Child Welfare Loans	P.O Box 25196 - 00603, Nairobi	Methodist Ministries Centre,Block C, Lavington, Nairobi	Nairobi
63	Parliamentarians Regulated Non-WDT-Sacco	For Members of Parliament and staff of the Parliamentary Service. This is a specialized SACCO serving the legislative arm of government and its administrative support structure.	BOSA Loans, Constituency Development Loans, Asset Finance	P.O Box 44809 - 00100, Nairobi	Protection House Building, Nairobi	Nairobi
64	PCEA Kayole Regulated Non-WDT-Sacco	For the congregation of the PCEA Kayole Parish. It represents a parish-level financial initiative under the broader umbrella of the Presbyterian Church of East Africa's development agenda.	BOSA Loans, Church Building Loans, Welfare & Benevolence	P.O Box 967 - 00518, Nairobi	PCEA Kayole Church, Spine Road, Nairobi	Nairobi

65	PCEA Regulated Non-WDT-Sacco	A national SACCO for members of the Presbyterian Church of East Africa. It operationalizes the church's social ministry through structured financial intermediation and member empowerment.	BOSA Loans, Church Project Financing, Clergy Vehicle Loans	P.O Box 27573 - 00506, Nairobi	PCEA Jitegemea House, Along Muhoho Avenue, South C, Nairobi	Nairobi
66	PEFA Nairobi Central Regulated Non-WDT-Sacco	For members of the PEFA (Pentecostal Evangelistic Fellowship of Africa) church in Nairobi Central. It facilitates economic fellowship and mutual support among a defined Pentecostal congregation.	BOSA Loans, Tithe Advance Loans, Ministry Support Loans	P.O Box 79459 - 00200, Nairobi	Solar House 3rd Floor, Harambee Avenue, Nairobi	Nairobi
67	PESA Regulated Non-WDT-Sacco	A SACCO focused on savings and investments ("Pesa" means Money). Its very name establishes a clear and direct value proposition centered on the accumulation and management of financial resources.	BOSA Loans, High-Yield Savings Plans, Investment Clubs	P.O Box 38622 - 00623, Nairobi	Mountain Mall, Along Thika Raod Nairobi	Nairobi
68	PICEA Staff Regulated Non-WDT-Sacco	For staff of the PICEA Group, potentially in the insurance sector (located at UAP Old Mutual Tower). It provides a dedicated financial services channel for employees of a specialized financial services firm.	BOSA Loans, Insurance-Linked Credit, Staff Development Loans	P.O Box 43013 - 00100, Nairobi	UAP - Old Mutual Tower, Upper Hill	Nairobi

69	Polytech Regulated Non-WDT-Sacco	For staff and students of the Technical University of Kenya. This SACCO promotes financial literacy and access to credit within an institution focused on applied sciences and technological education.	BOSA Loans, Technical Project Loans, Internship Allowance Advances	P.O Box 56679 - 00200, Nairobi	Technical University Of Kenya, Haile Selassie Avenue, Nairobi	Nairobi
70	Postbank Regulated Non-WDT-Sacco	For employees of the Postbank (now Consolidated Bank following merger). The SACCO has historically served staff of a key institution in promoting savings culture and rural financial inclusion.	BOSA Loans, Rural Outreach Financing, Mobile Banking Loans	P.O Box 30313 - 00100, Nairobi	Postbank House, Banda Street, Nairobi	Nairobi
71	Queensway Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. Its operations contribute to the dense network of financial cooperatives serving Nairobi's heterogeneous urban population.	BOSA Loans, Personal Emergency Loans, Sacco Shares Loans	P.O Box 48165 - 00100, Nairobi	Ojijo Plaza, Off Ojijo, Nairobi	Nairobi
72	Radio Guard Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base, sharing an address with NACICO. This collocation suggests potential administrative or operational synergies between distinct cooperative entities.	BOSA Loans, Security Equipment Loans, Guard Housing Loans	P.O Box 27546 - 00506, Nairobi	NACICO Plaza, Landhies Road, Nairobi	Nairobi

73	Rambhai Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. It operates as a member-owned financial intermediary, leveraging collective savings to provide accessible credit.	BOSA Loans, Small Business Loans, Jogoo Road Business Loans	P.O Box 41328 - 00500, Nairobi	Theevan House, Jogoo Road, City Stadium, Nairobi	Nairobi
74	Ramco Group Regulated Non-WDT-Sacco	For employees of the Ramco Group of companies. The SACCO is structured to meet the financial needs of staff within a conglomerate with diverse business interests.	BOSA Loans, Group-Wide Staff Loans, Corporate Investment Loans	P.O Box 18639 - 00500, Nairobi	KCB Building, Enterprise Road, Industrial Area	Nairobi
75	Reli Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. Its name, potentially derived from "reliability," underscores the fundamental cooperative value of dependable financial services for members.	BOSA Loans, Guaranteed Quick Loans, Reliable Emergency Cash	P.O Box 55541 - 00200, Nairobi	Reli Sacco, Workshop Road, Nairobi	Nairobi
76	Relief Regulated Non-WDT-Sacco	For staff of the International Committee of the Red Cross (ICRC) in Nairobi. It provides a financial community for personnel engaged in complex humanitarian operations in the region.	BOSA Loans, Emergency Deployment Loans, Humanitarian Project Advances	P.O Box 73226 - 00200, Nairobi	International Committee Of The Red Cross Offices, Nairobi	Nairobi

77	Royal Media Regulated Non-WDT-Sacco	For employees of Royal Media Services, a major TV and radio broadcaster. The SACCO serves as a financial pillar for professionals in the influential and fast-paced media industry.	BOSA Loans, Broadcast Equipment Loans, News Crew Travel Loans	P.O Box 7468 - 00300, Nairobi	Communication Centre, Maalim Juma Road, Nairobi	Nairobi
78	Rubani Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. It exemplifies the role of SACCOs in deepening financial inclusion through community-focused savings and loan products.	BOSA Loans, Jua Kali Sector Loans, Jitume Loans	P.O Box 57509 - 00200, Nairobi	Rubani House, Nairobi	Nairobi
79	Sauti Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base; "Sauti" means "Voice". The name metaphorically represents the principle of member democracy and participation in cooperative governance.	BOSA Loans, Member-Influenced Loan Products	P.O Box 30042 - 00100, Nairobi	Sauti Office, Kapiti Road, Nairobi	Nairobi
80	Sawa Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base; "Sawa" means "Okay" or "Good". The nomenclature communicates a commitment to fair and satisfactory financial services for all members.	BOSA Loans, Sawa Quick Loans, Fair Pricing Loans	P.O Box 46143 - 00100, Nairobi	WestPark Suites, Ojijo Road, Westlands	Nairobi
81	Shelloyees Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. Its operational framework is based on the mutual	BOSA Loans, Member-Guaranteed Loans,	P.O Box 73902 -	Vienna Court, State House	Nairobi

		guarantee principle, where members collectively underwrite the risk of the loan portfolio.	Solidarity Group Loans	00200, Nairobi	Crescent Road, Nairobi	
82	Smart Savers Regulated Non-WDT-Sacco	Likely serves employees of Isuzu East Africa, focusing on smart savings. The name explicitly promotes a strategic and disciplined approach to personal financial management and accumulation.	BOSA Loans, Targeted Savings Plans, Goal-Oriented Loans	P.O Box 30527 - 00100, Nairobi	Isuzu East Africa Premises, Nairobi	Nairobi
83	Stoke - UK Diaspora Regulated Non-WDT-Sacco	Caters to Kenyans in the diaspora living in Stoke-on-Trent or surrounding areas in the UK. It facilitates transnational financial flows and investment linkages with the Kenyan economy.	UK Property Loans, Diaspora Business Loans, Forex-Friendly Savings	P.O Box 73458 - 00200, Nairobi	Sinkeed Arcade,Kitengela,	Nairobi
84	Taa Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base; "Taa" means "Light". The name symbolizes a mission to provide financial enlightenment, guidance, and opportunity to its members.	BOSA Loans, Financial Literacy-Linked Loans, Business Startup Loans	P.O Box 10535 - 00100, Nairobi	Finance House, Nairobi CBD	Nairobi
85	CFAO Regulated Non-WDT-Sacco	For employees of CFAO Motors Kenya. The SACCO is embedded within the automotive sales and distribution sector, serving its workforce's unique financial needs.	BOSA Loans, Vehicle Purchase Loans, Spare Parts Financing	P.O Box 3391 - 00506, Nairobi	Yamaha Building, CFAO Motors Limited	Nairobi

					Premises Nairobi	
86	Tetra Pak Regulated Non-WDT-Sacco	For employees of Tetra Pak, the food packaging and processing company. It supports the financial well-being of staff in a multinational corporation specializing in packaging solutions.	BOSA Loans, Packaging Machine Loans, Tech Upgrade Loans	P.O Box 78340 - 00507, Nairobi	SMK Building, Industrial Area, Enterprise Road, Nairobi	Nairobi
87	The Standard Regulated Non-WDT-Sacco	For employees of the Standard Group media house. This internal SACCO underscores the prevalence of staff welfare associations within Kenya's major media corporations.	BOSA Loans, Media House Operations Loans, Staff Welfare Loans	P.O Box 30080 - 00100, Nairobi	The Standard Group Centre, Nairobi	Nairobi
88	Total Regulated Non-WDT-Sacco	For employees of TotalEnergies Marketing Kenya. The SACCO serves personnel in the downstream petroleum sector, a critical component of the national energy supply chain.	BOSA Loans, Fuel Station Operations Loans, LPG Distribution Loans	P.O Box 30675 - 00100, Nairobi	Regal Plaza, 2nd Floor West Wing, Nairobi	Nairobi
89	Transglob Regulated Non-WDT-Sacco	Likely serves the transport and logistics sector with a global focus. Its name indicates an orientation towards serving members with international or cross-border logistical operations.	BOSA Loans, International Transport Loans, Freight Financing	P.O Box 11364 - 00400, Nairobi	Acceler Towers, Nairobi	Nairobi

90	Uaminifu Regulated Non-WDT-Sacco	A SACCO operating from KWFT Centre; "Uaminifu" means "Honesty". This principle is foundational to building the trust necessary for the functioning of any financial intermediary.	BOSA Loans, Transparent Pricing Loans, Trust-Based Lending	P.O Box 4179 - 00506, Nairobi	KWFT Centre, Kiambere Road, Upper Hill	Nairobi
91	Ubora Regulated Non-WDT-Sacco	For employees of the Kenya Bureau of Standards (KEBS); "Ubora" means "Quality". The name reflects the institutional ethos of its sponsor organization and its commitment to high standards.	BOSA Loans, Quality Assurance Project Loans, Standards Compliance Loans	P.O Box 54974 - 00200, Nairobi	KEBS Headquarters, Popo Road, Nairobi	Nairobi
92	Ukaguzi Regulated Non-WDT-Sacco	Serves staff of the office of the Auditor-General ("Ukaguzi" means Audit). The SACCO supports the financial needs of professionals tasked with ensuring fiscal accountability in the public sector.	BOSA Loans, Audit Process Financing, Professional Indemnity Loans	P.O Box 61055 - 00200, Nairobi	Anniversary Towers, Nairobi	Nairobi
93	United Winners Regulated Non-WDT-Sacco	A community-based SACCO, likely in the Umoja area. Its aspirational name projects a vision of collective socioeconomic success and achievement for its membership.	BOSA Loans, Winners' Circle Loans, Success Project Financing	P.O Box 1390 - 00515, Westlands	Imani Towers, Umoja Estate, Nairobi	Nairobi

94	Unbound Regulated Non-WDT-Sacco	A community-based SACCO in the Kangemi area. The name suggests a mission to liberate members from financial constraints and unlock their economic potential.	BOSA Loans, Limitless Loans, Freedom Business Loans	P.O Box 1163 - 00100, Nairobi	Mountain view, House 97, Kangemi Area, Nairobi	Nairobi
95	Unga Regulated Non-WDT-Sacco	For employees of Unga Group, a major miller and food manufacturer. The SACCO is an integral part of the employee value proposition within a foundational agribusiness firm.	BOSA Loans, Milling Plant Loans, Food Fortification Loans	P.O Box 30386 - 00100, Nairobi	Ngano House, Industrial Area	Nairobi
96	United Women Regulated Non-WDT-Sacco	A SACCO focused on empowering women through financial inclusion. It addresses specific gender-based barriers to finance and promotes women's economic agency and entrepreneurship.	BOSA Loans, Women-Only Business Loans, Mama Biashara Loans	P.O Box 3228 - 00100, Nairobi	Green House Mall, Adams Arcade, Nairobi	Nairobi
97	Uokoaji Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base; "Uokoaji" means "Salvage" or "Rescue". The name implies a strategic focus on providing financial solutions during periods of economic distress.	BOSA Loans, Business Rescue Loans, Debt Relief Loans	P.O Box 30793 - 00100, Nairobi	Commodore OfficeSuites, Suite4c,Nairobi	Nairobi
98	USIU-Africa Regulated Non-WDT-Sacco	For staff and students of the United States International University-Africa. It fosters a culture of financial responsibility and planning within an international private university setting.	BOSA Loans, International Student Loans,	P.O Box 14634 - 00800, Nairobi	USIU, Maintenance Yard, Nairobi	Nairobi

			Faculty Research Loans			
99	Utabibu Regulated Non-WDT-Sacco	For medical doctors and professionals ("Utabibu" means medical treatment). This SACCO constitutes a professional financial collective for practitioners in the curative healthcare sector.	BOSA Loans, Medical Practice Startup Loans, Hospital Equipment Loans	P.O Box 102697 - 00101, Nairobi	Doctors' Park, 4th Floor, Parklands Avenue Nairobi	Nairobi
100	Utafiti Regulated Non-WDT-Sacco	For research scientists and staff of the International Livestock Research Institute (ILRI). It serves a community of researchers dedicated to agricultural innovation and livestock development.	BOSA Loans, Research Project Financing, Scientific Equipment Loans	P.O Box 30709 - 00100, Nairobi	ILRI Campus, Nairobi	Nairobi
101	Uwezo Regulated Non-WDT-Sacco	Likely serves employees of Crown Paints; "Uwezo" means "Capability". The name reflects an empowerment-oriented philosophy, aiming to enhance members' economic capacity and agency.	BOSA Loans, Capacity Building Loans, Empowerment Financing	P.O Box 78799 - 00507, Nairobi	Crown Paints, Likoni Road, Nairobi	Nairobi
102	Vegpro Regulated Non-WDT-Sacco	For employees of Vegpro Kenya, a major horticulture company. The SACCO is situated within the export-oriented horticulture sector, a	BOSA Loans, Horticulture Export Loans, Greenhouse Financing	P.O Box 3774 - 00100, Nairobi	Jomo Kenyatta International Airport-	Nairobi

		significant contributor to foreign exchange earnings.			Cargo Village, Nairobi	
103	Verona Huruma Regulated Non-WDT-Sacco	A church-based SACCO associated with the Holy Trinity Catholic Church in Kariobangi. It represents a grassroots faith-based initiative for community economic development.	BOSA Loans, Parish Project Loans, Huruma Community Loans	P.O Box 47714 - 00100, Nairobi	Holy Trinity Catholic Church, Kariobangi Area, Nairobi	Nairobi
104	Vision Regulated Non-WDT-Sacco	For staff of World Vision Kenya, the Christian humanitarian organization. The SACCO aligns with the organization's broader mission of human transformation and community development.	BOSA Loans, Community Vision Loans, Child Sponsorship Support Loans	P.O Box 1240 - 00502, Nairobi	World Vision Complex,KarenRoad, Nairobi	Nairobi
105	Wasado Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. Its operations are based on the core cooperative tenets of voluntary and open membership, democratic control, and member economic participation.	BOSA Loans, Member-Chosen Loan Products, Democratic Finance	P.O Box 18011 - 00500, Nairobi	KCB Building Industrial Area, 3rdFloor, Nairobi	Nairobi
106	Waskom Regulated Non-WDT-Sacco	A SACCO serving a diverse membership base. It functions as a member-owned, member-	BOSA Loans, Local Investment Loans,	P.O Box 17592 -	Blue House, Kangundo	Nairobi

		governed financial intermediary focused on local capital formation.	Wajukuu Wa Mtaa Loans	00500, Nairobi	Road, Nairobi	
107	Kenya - USA Diaspora Regulated Non-WDT-Sacco	Caters specifically to Kenyans living in the United States. It provides a structured platform for channeling diaspora investments into the Kenyan economy through a regulated entity.	US Dollar Denominated Loans, Stateside Property Loans, Cross-Border Investment	P.O Box 21372 - 00100, Nairobi	Crescent Business Centre, Parklands Area	Nairobi
108	Forward Travellers Regulated Non-WDT-Sacco	A community-based SACCO in the Kayole area. The name evokes a sense of progress and forward momentum in the financial journey of its members.	BOSA Loans, Journey Loans, Progressive Step Loans	P.O Box 1272 - 00518, Nairobi	Kayole Area, Nairobi	Nairobi

Appendix 5: University Introduction Letter

Kirinyaga University

Tel: +254 701562092, +254 728499650, +254 709742000/30
P.O. Box 143-10300 Kerugoya.

Email: info@kyu.ac.ke
Website: www.kyu.ac.ke

SCHOOL OF PURE AND APPLIED SCIENCES
DEPARTMENT OF PURE AND APPLIED SCIENCES

The Chief Executive Officer

National Commission for Science

Technology &Innovation (NACOSTI)

NAIROBI.

20th June, 2025.

Dear Sir/Madam

REF: RESEARCH PERMIT FOR RICHARD NDERITU MATHENGE PA303/S/21245/23

The above subject refers.

The above named is a student of Kirinyaga University pursuing a degree course in PhD in Information Technology in the Department of Pure and Applied Sciences. He has successfully defended his research proposal *"Adaptive GRU-based Neural network framework for Detecting and Mitigating phishing attacks in Digital microfinance ecosystems"* fulfillment for the degree award.

We are kindly requesting your office to provide him with the necessary assistance for data collection and completion of his research.

Any assistance offered to the student will be highly appreciated.

Please to do not hesitate to contact the undersigned for any information regarding the same.

Thank you.

KIRINYAGA UNIVERSITY
C.O.D. DEPARTMENT OF PURE
& APPLIED SCIENCES

Dr. Agatha Wagutu
Ag. COD, PURE & APPLIED SCIENCES

Appendix 6: NACOSTI Research Permit

REPUBLIC OF KENYA

NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Ref No: 569154

Date of Issue: 19/July/2025

RESEARCH LICENSE



This is to Certify that Mr.. Richard Mathenge of Kirinyaga University, has been licensed to conduct research as per the provision of the Science, Technology and Innovation Act, 2013 (Rev.2014) in Nairobi on the topic: Adaptive GRU-based Neural network framework for Detecting and Mitigating phishing attacks in Digital microfinance ecosystems for the period ending : 19/July/2026.

License No: NACOSTI/P/25/4176419

569154

Applicant Identification Number

Ag. Director General
NATIONAL COMMISSION FOR SCIENCE,TECHNOLOGY & INNOVATION

Verification QR Code

NOTE: This is a computer generated License. To verify the authenticity of this document, Scan the QR Code using QR scanner application.

See overleaf for conditions

Appendix 7: Code Repository

```
# ----------------------------
# Text cleaning helpers
# ----------------------------
def clean_text(text):
    if not isinstance(text, str):
        return ""
    text = text.lower()
    text = re.sub(r"http\S+|www\.\S+", " <url> ", text)
    text = re.sub(r"\b[\w\.-]+@[\w\.-]+\.\w+\b", " <email> ", text)
    text = re.sub(r"\d+", " <num> ", text)
    text = re.sub(r"[^a-zA-Z<>_'\s]", " ", text)
    text = re.sub(r"\s+", " ", text).strip()
    return text

TAG_RE = re.compile(r"<[^>]+>")
def html_to_text(s: str) -> str:
    s = TAG_RE.sub(" ", s or "")
    s = ihtml.unescape(s)
    s = re.sub(r"\s+", " ", s).strip()
    return s
```

```
# ----------------------------
# Email parsing (bytes -> dict)
# ----------------------------
def extract_email_content_bytes(raw: bytes, path_hint: str = "", truncate_chars: int = None):
    if not raw:
        return None
    try:
        msg = BytesParser(policy=policy.default).parsebytes(raw)
    except Exception:
        return None

    subject = msg.get('subject') or ""
    from_   = msg.get('from') or ""

    body = ""
    if msg.is_multipart():
        for part in msg.walk():
            if part.get_content_type() == "text/plain":
                try:
```

```
                body = part.get_content()
            except Exception:
                payload = part.get_payload(decode=True)
                body = payload.decode("utf-8", errors="ignore") if payload else ""
            if body:
                break
    if not body:
        for part in msg.walk():
            if part.get_content_type() == "text/html":
                try:
                    body = html_to_text(part.get_content())
                except Exception:
                    payload = part.get_payload(decode=True)
                    body = html_to_text(payload.decode("utf-8", errors="ignore")) if payload else ""
                break
else:
    ctype = msg.get_content_type()
    try:
        solo = msg.get_content()
    except Exception:
        payload = msg.get_payload(decode=True)
        solo = payload.decode("utf-8", errors="ignore") if payload else ""
    body = html_to_text(solo) if ctype == "text/html" else (solo or "")

clean_body    = clean_text(body) if body else ""
clean_subject = clean_text(subject) if subject else ""
full_text     = f"{clean_subject} {clean_body}".strip()

if truncate_chars and len(full_text) > truncate_chars:
    full_text = full_text[:truncate_chars]

if (subject and clean_subject) or (len(clean_body) >= 5):
    return {
        'subject': subject,
        'from': from_,
        'clean_text': full_text,
        'file_path': path_hint
    }
return None
```

```
# ----------------------------
# Labeling (applied per chunk)
# ----------------------------
def add_phishing_labels_inplace(rows: list):
```

```
    # add 'label' to each row dict
    indicators = {
        'password': 2, 'update.*password': 3, 'verify.*account': 3,
        'click.*here': 2, 'http://': 1, 'www\\.': 1, 'urgent': 2,
        'security.*alert': 3, 'account.*suspended': 3, 'confirm.*identity': 3,
    }
    pat_list = [(re.compile(p), w) for p, w in indicators.items()]
    for r in rows:
        text = r.get('clean_text', '') or ''
        subj = r.get('subject', '') or ''
        tl = text.lower()
        sl = subj.lower()
        score = 0
        for pat, w in pat_list:
            if pat.search(tl) or pat.search(sl):
                score += w
        r['label'] = 1 if score >= 3 else 0


# ----------------------------
# Model Architecture Functions
# ----------------------------
@tf.keras.utils.register_keras_serializable(package="custom")
class SimpleAttention(Layer):
    def __init__(self, **kwargs):
        super().__init__(**kwargs)
    def build(self, input_shape):
        hidden_dim = int(input_shape[-1])
        self.W = self.add_weight(name="att_weight", shape=(hidden_dim, 1),
                                 initializer="glorot_uniform", trainable=True)
        self.b = self.add_weight(name="att_bias", shape=(1,), initializer="zeros", trainable=True)
        super().build(input_shape)
    def call(self, inputs, mask=None):
        e = tf.tanh(tf.matmul(inputs, self.W) + self.b)
        if mask is not None:
            mask = tf.cast(mask, e.dtype)[:, :, tf.newaxis]
            neg_inf = tf.constant(-1e9, dtype=e.dtype)
            e = tf.where(tf.equal(mask, 1), e, neg_inf)
        a = tf.nn.softmax(e, axis=1)
        context = tf.reduce_sum(inputs * a, axis=1)
        return context
    def get_config(self):
        return super().get_config()
```

```
# ----------------------------
# Main execution (stream -> CSV -> load subset -> train)
# ----------------------------
os.makedirs(config['output_dir'], exist_ok=True)
csv_out = os.path.join(config['output_dir'], config['csv_out_name'])

# A) Decide data source
if config['data_csv']:
    print(f"Loading dataset from existing file: {config['data_csv']}")
    data_path = config['data_csv']
else:
    # Stream from maildir, fallback to archive if needed
    try:
        print(f"No CSV provided. Streaming maildir: {config['maildir_root']}")
        written = stream_maildir_to_csv(
            root_dir=config['maildir_root'],
            csv_path=csv_out,
            rows_per_flush=config['rows_per_flush'],
            parse_max=config['parse_max_emails'],
            truncate_chars=config['truncate_chars'],
        )
        data_path = csv_out
    except Exception as e:
        print(f"[Info] Maildir streaming failed: {e}")
        if config.get('archive_path'):
            print(f"[Info] Falling back to archive: {config['archive_path']}")
            written = stream_archive_to_csv(
                archive_path=config['archive_path'],
                csv_path=csv_out,
                rows_per_flush=config['rows_per_flush'],
                parse_max=config['parse_max_emails'],
                truncate_chars=config['truncate_chars'],
                subdir_hint='maildir'
            )
            data_path = csv_out
        else:
            raise RuntimeError(
                "Maildir failed and no archive fallback configured. "
                "Set config['archive_path'] to your enron_mail_20150507.tar(.gz)."
            ) from e

# B) Load a bounded subset for training to keep RAM stable
print(f"\nLoading up to {config['train_rows']} rows for training from: {data_path}")
df = pd.read_csv(data_path, nrows=config['train_rows'])
```

Appendix 8: Certificate of Acceptance (JCS)



CERTIFICATE OF ACCEPTANCE

Certificate of acceptance for the paper (ID: 77183) titled:
Phishing, Vulnerabilities, and AI Defense: A Systematic Review of Cybersecurity Challenges and GRU-Based Mitigation Strategies in Digital Microfinance Institutions

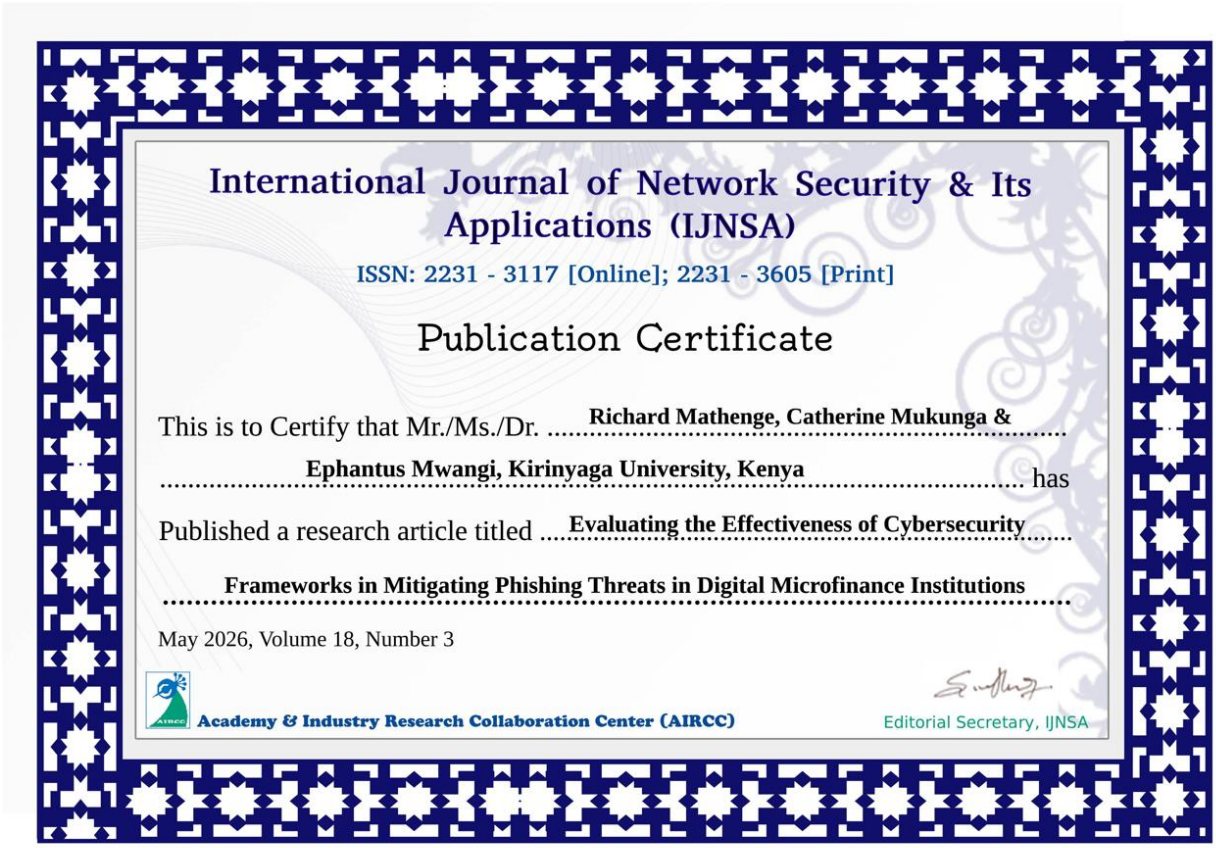
Authored by:
Richard Mathenge, Catherine Mukunga,
Ephantus Mwangi

has been accepted in Journal of Cyber Security
on 27 January 2026

Journal of
Cyber Security

Tech Science Press

Appendix 9: Certificate of Acceptance (IJNSA)



International Journal of Network Security & Its Applications (IJNSA)

ISSN: 2231 - 3117 [Online]; 2231 - 3605 [Print]

Publication Certificate

This is to Certify that Mr./Ms./Dr. **Richard Mathenge, Catherine Mukunga & Ephantus Mwangi, Kirinyaga University, Kenya** has

Published a research article titled **Evaluating the Effectiveness of Cybersecurity Frameworks in Mitigating Phishing Threats in Digital Microfinance Institutions**

May 2026, Volume 18, Number 3

Academy & Industry Research Collaboration Center (AIRCC)

Editorial Secretary, IJNSA