AN ANALYTICAL STUDY OF AI-ENHANCED SOCIAL ENGINEERING ATTACKS AND THEIR IMPACT ON MODERN CYBERSECURITY: A CASE STUDY OF THE UNIVERSITY OF BENIN (UNIBEN), NIGERIA

BY

CLARISSA EJEMAI OLOHIRENUAN

PSC2105324

DEPARTMENT OF COMPUTER SCIENCE
FACULTY OF PHYSICAL SCIENCES
UNIVERSITY OF BENIN

NOVEMBER 2025

CERTIFICATION

This is to certify that **CLARISSA EJEMAI OLOHIRENUAN**, with Matriculation number **PSC2105324**, carried out this project work under my supervision. It is adequate and satisfactory, both in scope and content, for the award of Bachelor of Science (B.Sc.) Degree in Computer Science from the University of Benin.

_____ _____

MR. IMIEFOH, P.E.B. **DATE**

Project Supervisor

APPROVAL

This project is hereby approved in partial fulfilment of the requirements for the award of Bachelor of Science (B.Sc.) Degree in Computer Science from the University of Benin.

_____ _____

DR. ROSEMARY USIOBIAFO DATE

Head of Department

DEDICATION

This work is dedicated first and foremost to GOD almighty who equipped me with the strength to complete this project. I also dedicate this work to my family and friends, whose unwavering support and encouragement have guided me throughout this research journey. And lastly, I dedicate it to all students and cybersecurity practitioners striving to create safer digital spaces.

ACKNOWLEDGEMENT

First and foremost, I thank God Almighty for His endless grace, wisdom, and strength, without which this journey would not have been possible. His guidance has carried me through every challenge, doubt, and milestone.

I am deeply grateful to my loving parents, Mr. and Mrs. Ejemai, and my wonderful siblings, Gianna, Colette, and Michelle. Your unwavering love, encouragement, and belief in me have been my anchor and my inspiration.

To my uncle, Mr. Osagie Edobor, thank you for your constant support, both seen and unseen you have been a pillar of strength throughout my academic journey

I am also profoundly thankful to my aunt, Mrs. Omoye Osayi Edobor, whose care, encouragement, and guidance have been a source of comfort and motivation.

To my relatives, the Odions, my aunt Laura Ejemai, and my cousin Ella Adeyeye-Ejemai, thank you for your immeasurable love, support, and faith in my abilities. Your encouragement has lifted me in moments of doubt and inspired me to persevere.

I also owe a heartfelt thanks to my friends, roommates, and coursemates, Emmy, Nehita, Eunice, Christabel, Joanna, Joshua Idogho, Jude Onose, Aneesa, Gabriel, Pascal, and Charles. Your friendship, encouragement, and endless support have made this journey lighter, brighter, and more memorable.

I would also like to express my heartfelt appreciation to the family I found at the Catholic Charismatic Renewal Students Community in Nigeria, St Albert Parish, University of Benin Ugbowo. Thank you for giving me a platform to grow, for always showing up, and for nurturing me spiritually, emotionally, and socially throughout this journey.

Finally, I am grateful to my project supervisor, Mr. Imiefo, for his guidance, patience, and unwavering belief in my potential

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

The rapid advancement of artificial intelligence (AI) has transformed cybersecurity, enabling attackers to use AI-enhanced social engineering techniques to exploit human behavior and bypass traditional defenses. This study investigates the impact of AI-driven social engineering attacks on modern cybersecurity, using the University of Benin (UNIBEN) student population as a case study. A survey was administered to 140 respondents to assess awareness, exposure, and perceptions of AI-powered scams, including phishing, deepfake impersonation, voice cloning, and chatbot-based attacks. The findings indicate that a majority of students (95%) are aware of social media scams, with 56% reporting prior victimization. Fake social media accounts, WhatsApp/Telegram scams, and phishing emails were the most commonly encountered threats. Participants reported difficulty distinguishing AI-generated attacks from human-driven scams, highlighting a significant vulnerability. The study further identifies gaps in current cybersecurity measures and emphasizes the need for enhanced awareness, training, and AI-aware defense frameworks. The results contribute to understanding the evolving threat landscape of AI-enhanced social engineering and provide actionable insights for improving digital security in academic institutions and beyond.

CHAPTER ONE

INTRODUCTION

1.0 Background to the Study

The rapid advancement of Artificial Intelligence (AI) has significantly transformed the field of cybersecurity, influencing both defensive and offensive operations. AI-driven tools have enhanced the detection and prevention of cyber threats, while adversaries have begun to exploit the same technology to develop more complex and deceptive attack strategies. One of the most critical developments emerging from this trend is the rise of AI-enhanced social engineering attacks, which combine intelligent automation, personalization, and manipulation of human psychology to achieve malicious objectives (Bharati, 2024).

Traditional social engineering attacks often depend on human deception through methods such as phishing emails, pretexting, and baiting. However, the integration of AI technologies has made these attacks more sophisticated and difficult to identify. Attackers can now use AI to create realistic fake messages, deepfake videos, synthetic voices, and automated chatbot interactions that convincingly imitate legitimate communication (Falade, 2023; Schmitt & Flechais, 2023). For instance, AI-powered tools such as FraudGPT and WormGPT have been reported to facilitate large-scale phishing and business email compromise (BEC) attacks with high levels of precision and automation (Falade, 2023).

Globally, the scale of AI-driven cyber threats continues to increase. According to recent cybersecurity reports, phishing remains responsible for more than 36 percent of data breaches, and this percentage is expected to rise as generative AI tools become more accessible (Sabatini, 2025). The World Economic Forum (2024) also identifies AI-enabled cybercrime as one of the top five threats to digital security worldwide, with projected economic losses estimated at 10.5 trillion United States dollars annually by 2025 (Bharati, 2024).

Despite these developments, existing research continues to focus primarily on traditional social engineering approaches, thereby neglecting the evolving nature and impact of AI-enhanced techniques. This has created a significant research gap, particularly in developing countries such as Nigeria, where technological adoption is increasing rapidly but awareness of AI-related cyber risks remains limited.

This study, therefore, aims to analyze the evolving tactics, vulnerabilities, and implications of AI-enhanced social engineering attacks within the context of modern cybersecurity. The study focuses on the University of Benin (UNIBEN) as a case study to assess the level of awareness, perception, and preparedness of individuals in responding to AI-driven social engineering threats.

1.1 Statement of the Problem

Social engineering attacks remain among the most effective techniques for compromising cybersecurity systems because they exploit human behavior rather than technical flaws. With the emergence of Artificial Intelligence (AI), particularly generative AI technologies, the complexity and reach of such attacks have increased considerably. Modern cybercriminals now utilize AI-powered tools such as ChatGPT-based variants, including FraudGPT and WormGPT, to craft convincing phishing messages, generate deepfake voices, and deploy automated malicious chatbots capable of deceiving even trained users (Falade, 2023; Schmitt & Flechais, 2023).

Conventional security measures such as employee awareness programs, two-factor authentication, and rule-based detection systems are increasingly insufficient to counter these threats (Bharati, 2024; Sabatini, 2025). This limitation arises from the ability of AI systems to personalize attacks on a large scale, automate psychological manipulation, and evade traditional security filters (Gupta et al., 2023). Additionally, the human-like fluency of AI-generated content and the rapid evolution of generative models make detection and response even more challenging (Almutairi & Elgibreen, 2022).

The absence of comprehensive, AI-aware cybersecurity frameworks has created a critical vulnerability, particularly in developing contexts where awareness and preparedness remain low. Although several studies have examined social engineering and AI independently, few have

analyzed their convergence and its implications for cybersecurity practices (Kumar & Patel, 2025; Rohini et al., 2025).

In the Nigerian context, universities and research institutions such as the University of Benin (UNIBEN) are becoming increasingly reliant on digital communication systems, making them potential targets for AI-driven deception and manipulation. Without a deep understanding of how AI-enhanced social engineering operates and affects cybersecurity systems, these institutions remain at risk of data breaches, identity theft, and reputational harm.

This study, therefore, seeks to bridge this gap by examining the dynamics of AI-enhanced social engineering attacks, their specific manifestations within the University of Benin (UNIBEN) environment, and potential strategies to strengthen institutional cybersecurity defenses.

1.2 Research Aim and Objectives

1.2.1 Aim of the Study

The main aim of this study is to analyze and evaluate the impact of Artificial Intelligence (AI)-enhanced social engineering attacks on modern cybersecurity frameworks, using the University of Benin (UNIBEN) as a case study. It also aims to assess how prepared individuals and institutions are to respond to these threats, and to develop practical strategies and frameworks that can strengthen cybersecurity awareness and defense within educational settings like UNIBEN.

1.2.2 Specific Objectives

To achieve the aim, the study sets out the following specific objectives:

1. To examine the evolving tactics and techniques used in AI-enhanced social engineering attacks, highlighting how AI has improved the precision, personalization, and success rate of these cyber threats.
2. To analyze the role of generative AI tools such as ChatGPT, FraudGPT, and WormGPT in creating convincing phishing messages, fake identities, and other deceptive content used to trick victims.

3. To investigate the psychological and institutional vulnerabilities that AI-powered attackers exploit, focusing on how human behavior, lack of awareness, and weak cybersecurity practices contribute to the success of social engineering attacks.
4. To assess the level of awareness and effectiveness of existing cybersecurity defense strategies within the University of Benin (UNIBEN), identifying areas where current systems or policies may not adequately address AI-driven threats.
5. To propose a practical framework or set of recommendations that can help mitigate AI-enhanced social engineering attacks by improving awareness, training, and the adoption of AI-assisted defense mechanisms within UNIBEN and similar institutions.

1.3 Research Questions

This study seeks to answer the following research questions:

1. What are the emerging tactics and techniques used in AI-enhanced social engineering attacks?
2. How do generative AI tools such as ChatGPT, FraudGPT, and WormGPT contribute to the sophistication and success of social engineering campaigns?
3. What cognitive and technical vulnerabilities do AI-powered attackers exploit to compromise information systems and manipulate human behavior?
4. To what extent are existing cybersecurity measures at the University of Benin (UNIBEN) effective in detecting and mitigating AI-driven social engineering threats?
5. What practical strategies or frameworks can be developed to strengthen institutional resilience against AI-enhanced social engineering attacks?

1.4 Significance of the Study

The relevance of this study lies in its contribution to understanding and addressing one of the fastest-evolving threats in cybersecurity AI-enhanced social engineering. The study offers value across several dimensions:

1. Academic Contribution:

This research adds to the growing body of literature at the intersection of artificial intelligence

and cybersecurity. Analysing the new forms of deception created through AI tools, it provides a foundation for future scholarly inquiry into how technology can both threaten and secure human communication systems (Bharati, 2024; Falade, 2023).

2. Practical Relevance for Cybersecurity Professionals:
Cybersecurity practitioners and institutional IT administrators will benefit from a clearer understanding of how generative AI tools are weaponized in phishing, impersonation, and manipulation. The findings may help shape more adaptive defense mechanisms suitable for academic environments like UNIBEN, where digital communication and data sharing are routine (Sabatini, 2025).

3. Policy and Regulatory Implications:
The study's insights may assist policymakers and university authorities in formulating updated cybersecurity policies, ethical AI guidelines, and data protection frameworks to address the rising threats of AI-enabled deception (Gupta et al., 2023).

4. Enhanced Human Awareness and Training:
By identifying the psychological and behavioural weaknesses that AI-based attacks exploit, this study can inform targeted awareness campaigns and capacity-building initiatives for staff and students within the university community (Kaur, 2025).

In summary, this study aims to bridge the gap between rapid AI innovation and the need for stronger, human-centred cybersecurity defences within academic institutions in Nigeria.

1.5 Scope and Limitations of the Study

Scope:
This study focuses on examining how artificial intelligence enhances social engineering attacks and how such attacks impact cybersecurity frameworks within academic institutions, using the University of Benin (UNIBEN) as a case study. It investigates key AI-driven attack types, including phishing, spear-phishing, voice cloning, chatbots, and deepfake-based impersonation. The research further explores the human and technical factors that contribute to system vulnerability, as well as institutional strategies that can be developed to improve awareness and

resilience. The temporal focus covers the period from 2020 to 2025, corresponding to the widespread adoption and misuse of generative AI tools in cybersecurity contexts.

Limitations:

While this research provides valuable insights into AI-enhanced social engineering attacks within the University of Benin community, certain limitations must be acknowledged:

1. **Sample Size Constraint:** The sample size of 140 respondents, though sufficient for descriptive analysis, may not fully represent the entire university population of approximately 60,000 students. This limits the generalizability of the findings across all faculties, levels, and demographic groups.
2. **Self-Reported Data Bias:** The survey relied on self-reported responses, which are inherently subject to bias. Some participants may have exaggerated their awareness levels or underreported scam experiences due to embarrassment, fear, or misinterpretation of questions.
3. **Restricted Participant Scope:** The study focused exclusively on students, thereby excluding academic and administrative staff who also interact frequently with digital platforms and may face different categories of cyber threats. Including these groups could have provided a more comprehensive understanding of AI-related social engineering within the institution.
4. **Rapid Technological Evolution:** The fast-paced evolution of AI technologies presents a dynamic challenge. Tools like deepfakes, voice cloning, and generative chatbots evolve rapidly, which means the perceptions captured in this study may change over time as both users and attackers adapt.
5. **Geographical Limitation**: The study was conducted within a single institution (University of Benin), which may limit the generalizability of the results to other universities or regions with different digital cultures, cybersecurity policies, or technological infrastructures.
6. **Rapid Evolution of AI Threats:** The dynamic nature of AI technologies presents an additional limitation. Emerging tools such as deepfakes, AI voice cloning, and generative

phishing systems evolve rapidly, potentially rendering current perceptions and awareness levels outdated within a short period

1.6 Definition of Terms

1. **Artificial Intelligence (AI):** A field of computer science that focuses on creating systems capable of performing tasks that require human intelligence, including reasoning, learning, and decision-making (Blauth, Gstrein, & Zwitter, 2022).
2. **Social Engineering:** A psychological manipulation technique that exploits human behavior and trust to gain unauthorized access to confidential information or systems (Rohini, Gomathi, & Others, 2025).
3. **AI-Enhanced Social Engineering:** The use of artificial intelligence technologies, such as machine learning and generative models, to automate and personalize social engineering attacks, increasing their success rate (Bharati, 2024).
4. **Generative AI:** A category of AI systems that can produce new content such as text, audio, or images based on patterns learned from large datasets. Examples include ChatGPT and similar models (Schmitt & Flechais, 2023).
5. **ChatGPT:** A large language model developed by OpenAI, designed to generate human-like text that can be used for conversation, content creation, and automation (Gupta et al., 2023).
6. **FraudGPT / WormGPT:** Malicious versions of generative AI models used by cybercriminals to create phishing messages, automate scams, and evade traditional security measures (Falade, 2023).
7. **Phishing:** A cyberattack method in which attackers impersonate legitimate entities to trick users into sharing sensitive information, often through emails or social media platforms (Kaur, 2025).
8. **Deepfake:** Digitally altered or synthetically generated media created using AI to imitate a person's face, voice, or actions, often used in impersonation or misinformation (Almutairi & Elgibreen, 2022).
9. **Threat Landscape:** The collective environment of existing and emerging cybersecurity threats, including the tactics, tools, and actors involved in cyberattacks (Sabatini, 2025).

10. **Natural Language Processing (NLP):** An area of AI focused on enabling machines to understand and generate human language (Bharati, 2024).
11. **Zero-Day Attack:** An attack that exploits unknown software vulnerabilities before a fix or patch is available, often detected through AI-enhanced reconnaissance (Kumar & Patel, 2025).
12. **Automated Chatbots:** AI-driven conversational agents that communicate with users via messaging or web platforms. These can be weaponized for phishing or impersonation (Ariza et al., 2023).
13. **Attack Surface:** The total number of potential points where an unauthorized user can try to access a system or extract data, which often increases with AI integration (Schmitt & Flechais, 2023).

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

Artificial intelligence (AI) has become a key driver of change in cybersecurity, influencing both defensive strategies and the ways attackers exploit human and system vulnerabilities. In particular, AI has transformed social engineering attacks, enabling cybercriminals to manipulate individuals more effectively and at a larger scale. Literature on this topic covers several areas, including the technical mechanisms behind AI-enhanced attacks, the psychological strategies used to deceive victims, and the methods organizations use to defend against such threats. This review synthesizes research on these areas to highlight key trends, challenges, and gaps that inform this study.

A growing body of research emphasizes the role of generative AI in increasing the sophistication of social engineering attacks. Studies by Bharati (2024) and Falade (2023) show how AI tools such as ChatGPT, FraudGPT, and WormGPT are used to create convincing phishing emails, spear-phishing campaigns, and deepfake-enabled impersonations. Similarly, Blauth, Gstrein, and Zwitter (2022) and Sabatini (2025) discuss the broader misuse of AI in cybercrime, illustrating the systemic risks these technologies pose to organizations and individuals.

On the defensive side, research highlights the development of AI-powered security solutions. Gomathi Alias Rohini et al. (2025) and Gupta et al. (2023) describe machine learning-based detection systems that identify suspicious behavior or communications. However, these studies also note significant limitations, as AI-enabled attacks often bypass conventional filters and exploit human judgment. Kumar and Patel (2025) and Schmitt and Flechais (2023) further emphasize the importance of integrating human oversight, or human-in-the-loop strategies, to strengthen cybersecurity against AI-driven social engineering.

This literature review focuses on three main areas:

1. The evolution of AI-enhanced social engineering techniques.
2. The impact of these attacks on organizational security and user privacy, with attention to the Nigerian academic environment, particularly the University of Benin (UNIBEN).
3. The effectiveness and limitations of existing defense mechanisms, including technological and human-centered strategies.

By analyzing these areas, the review identifies research gaps in regulation, ethical governance, and proactive mitigation strategies (Ariza et al., 2023; Almutairi & Elgibreen, 2022; Kaur, 2025). This synthesis provides the foundation for the current study, which seeks to explore AI-enhanced social engineering threats within the UNIBEN community and propose practical measures to enhance cybersecurity awareness and resilience.

2.1 Social Engineering

Social engineering in cybersecurity refers to the deliberate manipulation of human behavior to gain unauthorized access to information, systems, or physical resources. Unlike technical attacks that exploit system vulnerabilities, social engineering targets people the most unpredictable and often weakest link in security frameworks (Gomathi Alias Rohini et al., 2025; Kaur, 2025). By exploiting trust, curiosity, or fear, attackers can bypass advanced technological defenses with relative ease. This highlights that cybersecurity is not only a technical problem but also a socio-technical challenge where human behavior plays a central role.

The effectiveness of social engineering relies on well-known psychological principles such as authority, scarcity, reciprocity, and urgency. Attackers use these principles to create a sense of legitimacy or pressure, prompting individuals to act against their usual judgment. For example, phishing emails often impersonate banks, government agencies, or university authorities to elicit immediate responses, while pretexting relies on fabricated scenarios to extract sensitive data. These tactics exploit cognitive biases and decision-making shortcuts inherent in human behavior (Gomathi Alias Rohini et al., 2025).

Historically, social engineering involved simple methods such as phone-based pretexting, dumpster diving for confidential documents, or baiting with infected USB drives. Early attacks were limited by scale and personalization; attackers could reach only a finite number of targets,

often using generic templates, which made detection easier (Kaur, 2025). Despite these limitations, early social engineering attacks were highly successful because they directly exploited human trust and error, bypassing technological safeguards.

In today's hyperconnected digital environment, the significance of social engineering has increased. The adoption of digital tools, remote work, and the expansion of online networks have created more opportunities for attackers. High-profile data breaches frequently originate from human error induced through social engineering, which often serves as the first step in complex cyberattacks (Kumar & Patel, 2025; Nishant & Patel, 2025). Understanding these mechanisms is crucial for analyzing how artificial intelligence further amplifies the effectiveness and scale of social engineering attacks, particularly in academic institutions such as the University of Benin (UNIBEN).

2.2 Evolution of AI in Cyberattacks

The integration of artificial intelligence (AI) into cyber operations has transformed attacker capabilities. Cybercriminals have shifted from manual, script-based attacks to adaptive, data-driven systems capable of producing realistic text, audio, and multimedia at scale. This section reviews this evolution and highlights its implications for social engineering, particularly in academic institutions such as the University of Benin (UNIBEN).

2.2.1 Early Automation versus Modern AI

Early cyber tools focused on automation rather than sophistication. Spam engines sent mass emails, simple scripts collected online data, and basic botnets executed repetitive tasks. While these tools increased reach, they offered little personalization and relied on volume rather than effectiveness (Blauth et al., 2022). Attackers depended on static templates that were relatively easy for signature-based defenses to detect and block (Ariza et al., 2023).

The introduction of machine learning (ML) marked a significant change. ML systems analyze data patterns and adapt their actions based on observed behavior. Initially, ML assisted with reconnaissance and target selection. Contemporary systems now use natural language processing (NLP) to produce coherent, contextually accurate text (Bharati, 2024). This shift allows attackers

to conduct smaller, highly targeted campaigns that prioritize plausibility over volume (Gupta et al., 2023). As a result, defenders must go beyond static rules and develop systems capable of interpreting behavior and semantics (Schmitt & Flechais, 2023). Early automation, therefore, served as a stepping stone to an era where AI fundamentally changes attack strategies and resource allocation (Blauth et al., 2022; Ariza et al., 2023).

2.2.2 The Rise of Generative AI

Generative AI models, including large language models (LLMs), generative adversarial networks (GANs), and diffusion models, represent a disruptive leap in cyberattack capabilities. LLMs can generate fluent, contextually appropriate text, enabling adversaries to craft spear-phishing messages that appear legitimate (Schmitt & Flechais, 2023; Falade, 2023). GANs and diffusion models allow attackers to create realistic images, videos, and audio, producing convincing deepfakes and voice clones (Almutairi & Elgibreen, 2022).

Two key properties make generative AI particularly attractive to attackers: first, the quality of output is human-like enough to reduce suspicion; second, pretrained models and APIs make advanced capabilities accessible to non-experts (Gupta et al., 2023; Bharati, 2024). Malicious variants such as FraudGPT and WormGPT allow even unskilled actors to launch sophisticated attacks (Falade, 2023). Modern systems often combine multiple modalities text, voice, and video to increase credibility, creating challenges for current detection methods that typically focus on a single modality (Schmitt & Flechais, 2023; Almutairi & Elgibreen, 2022).

2.2.3 AI in Personalization and Automation

AI enables unprecedented levels of personalization and automation in attacks. Using publicly available data from social media, professional profiles, and organizational charts, attackers can craft messages tailored to individual targets (Bharati, 2024; Sabatini, 2025). Feedback loops allow attackers to adapt messages in real time based on responses, increasing the likelihood of success (Falade, 2023).

Studies indicate that hyper-personalized, AI-generated messages achieve higher engagement rates than generic phishing attempts, though precise metrics are limited (Ariza et al., 2023; Gupta

et al., 2023). Conversational AI tools such as chatbots maintain ongoing interactions, build trust, and escalate manipulative tactics when necessary (Ariza et al., 2023; Kumar & Patel, 2025). Voice cloning and real-time audio synthesis also extend social engineering to phone calls and live video, enabling highly impactful fraud (Almutairi & Elgibreen, 2022; Schmitt & Flechais, 2023).

AI's dual role reducing skill requirements for attackers and enabling continuous, adaptive campaigns creates an asymmetric risk environment. Attackers can scale personalized deception at low cost, while defenders must invest heavily in detection, training, and governance. Critical gaps include limited datasets for benchmarking AI-enhanced attacks, insufficient cross-modal detection research, and a lack of standardized methods to attribute AI involvement in incidents (Sabatini, 2025; Falade, 2023).

2.2.4 Synthesis

The literature consistently shows that AI has transformed cyberattacks from high-volume, low-specificity strategies to lower-volume, high-plausibility campaigns that exploit human cognition more effectively. Generative models have been central to this shift, improving both the quality and accessibility of deceptive content. Automation and personalization have operationalized attacks at scale, while detection and mitigation remain challenging (Bharati, 2024; Falade, 2023; Schmitt & Flechais, 2023). Key research priorities include developing cross-modal detection systems, creating shareable datasets for evaluation, and designing governance mechanisms to mitigate dual-use risks without hindering legitimate AI applications.

2.3 AI-Enhanced Social Engineering Techniques

Artificial intelligence has significantly transformed traditional social engineering methods by introducing automation, personalization, and multimodal deception. Unlike conventional attacks, which relied on manual effort and generic scripts, AI-powered social engineering can adapt to individual targets, mimic human communication patterns, and operate at scale. These developments increase the likelihood of successful attacks while challenging traditional cybersecurity defenses (Bharati, 2024; Falade, 2023; Schmitt & Flechais, 2023).

This section examines the main techniques through which AI enhances social engineering attacks. It covers AI-driven phishing, deepfake-based impersonation, voice cloning and vishing, and conversational AI attacks. Each subsection explores how attackers leverage AI to manipulate human psychology, automate campaigns, and bypass existing security measures. Emphasis is placed on both technical mechanisms and real-world implications, providing a foundation for understanding the growing threat landscape, particularly within organizational and academic settings such as the University of Benin (Gupta et al., 2023; Kaur, 2025; Sabatini, 2025).

2.3.1 AI-driven Phishing

Phishing remains one of the most prevalent social engineering techniques, but the integration of AI particularly large language models (LLMs) has revolutionized its scale, sophistication, and effectiveness. Traditionally, phishing attacks relied on generic, poorly written emails that could be easily detected by spam filters or identified by vigilant users. AI-driven phishing, however, introduces personalization, linguistic accuracy, and contextual relevance, significantly increasing success rates (Bharati, 2024; Falade, 2023).

2.3.1.1 Personalized Spear-Phishing Emails Using LLMs

Generative AI models such as GPT-based systems enable attackers to craft highly convincing and context-aware phishing emails. Unlike traditional scripts, LLMs can analyze public data, including social media profiles and corporate information, to personalize emails to individual targets (Gupta et al., 2023). This personalization creates a sense of legitimacy and trust, making it more difficult for users to detect fraudulent intent (Kumar & Patel, 2025). Furthermore, AI allows dynamic language adaptation matching tone, style, and cultural context which enhances believability and bypasses conventional security filters that rely on static keyword detection (Schmitt & Flechais, 2023).

2.3.1.2 Statistical Evidence of Improved Success Rates

Studies reveal a substantial increase in phishing effectiveness with AI integration. According to Falade (2023), AI-enhanced phishing emails achieved success rates up to 40% higher than

traditional phishing attempts, primarily due to superior linguistic accuracy and contextual cues. Bharati (2024) further notes that attackers can generate hundreds of unique, grammatically correct phishing emails per minute, making detection through signature-based methods nearly impossible. This scalability amplifies the threat landscape, enabling mass-targeted and spear-phishing campaigns simultaneously.

2.3.1.3 Examples and Real-World Implications

Recent research highlights AI-driven phishing being used in high-profile data breaches, where attackers utilized automated scripts to impersonate vendors and partners convincingly (Gupta et al., 2023). Additionally, phishing-as-a-service (PhaaS) platforms now integrate AI tools, lowering the technical barrier for cybercriminals and democratizing access to sophisticated attack techniques (Sabatini, 2025). These developments underscore the urgent need for adaptive security measures and AI-powered defenses.

2.3.2 Deepfake-based Impersonation

Deepfake technology, powered by Generative Adversarial Networks (GANs) and advanced AI models, has significantly amplified the threat of impersonation in social engineering attacks. Unlike traditional spoofing techniques that relied on static images or text-based deception, deepfakes allow attackers to produce highly realistic audio-visual content that convincingly imitates real individuals. This capability enables large-scale impersonation in scenarios such as CEO fraud, political disinformation, and identity-based scams.

According to Schmitt and Flechais (2023), deepfakes represent a paradigm shift in trust exploitation because they bypass traditional verification cues like facial expressions and voice tone, which humans typically rely on for authenticity. AI-driven deepfake generation tools now require minimal input often just a few seconds of audio or video making the barrier to entry for cybercriminals extremely low (Kumar & Patel, 2025).

High-profile incidents illustrate the severity of this threat. For example, voice deepfakes have been used to trick executives into authorizing fraudulent fund transfers, causing millions in losses. In one notable case, a deepfake audio impersonation of a company CEO successfully convinced a UK-based energy firm to transfer €220,000 to a scammer's account (Bharati, 2024).

Similarly, Sabatini (2025) predicts that deepfakes will become a standard tool in Business Email Compromise (BEC) and corporate scams, enabling real-time impersonation during video calls and conferences.

Furthermore, the social impact of deepfake impersonation extends beyond financial fraud. Disinformation campaigns leveraging synthetic media have targeted political figures, celebrities, and even ordinary individuals, damaging reputations and influencing public opinion (Schmitt & Flechais, 2023). These attacks are particularly effective because humans are hardwired to trust visual and auditory signals, making AI-driven impersonation far more persuasive than text-based deception alone.

2.3.3 Voice Cloning and Vishing

Voice cloning technology, powered by advanced AI models such as text-to-speech (TTS) systems and generative adversarial networks (GANs), has introduced a new dimension to vishing (voice phishing) attacks. Unlike traditional phone scams, where human impersonators relied on linguistic skills and psychological manipulation, AI-based voice synthesis enables attackers to generate highly realistic voices that mimic specific individuals with minimal input data. Studies indicate that only a few seconds of recorded speech are sufficient to create convincing voice clones (Schmitt & Flechais, 2023).

This capability has been exploited in high-profile incidents, such as the 2019 case in which cybercriminals used AI-generated speech to impersonate the CEO of a German company and fraudulently request a $243,000 transfer (Bharati, 2024). These attacks are particularly dangerous because they leverage two psychological principles: authority bias (trust in a perceived superior) and urgency, making targets more likely to comply without verification. Furthermore, the scalability of voice cloning means attackers can automate hundreds of calls simultaneously, dramatically increasing the success rate.

Recent research highlights a concerning trend: voice-based authentication systems, once considered secure, are now vulnerable to synthetic voice spoofing (Almutairi & Elgibreen, 2022). Despite efforts to develop anti-spoofing mechanisms, such as detecting artifacts in generated audio, adversaries continue to refine cloning models to bypass detection. The growing

sophistication of real-time voice synthesis poses an escalating threat to sectors reliant on verbal confirmations, such as banking, law enforcement, and executive-level communication.

Key Implications:

1. Low barrier to entry: Open-source voice synthesis tools reduce the technical expertise required for such attacks.
2. High impact on trust-based systems: Sectors relying on verbal confirmation (e.g., finance) face significant risks.
3. Detection challenges: Current voice biometrics and anti-spoofing technologies are lagging behind evolving cloning techniques.

2.3.4 Chatbot and Conversational Attacks

Conversational AI systems have emerged as a double-edged sword in the cybersecurity landscape. While legitimate applications of chatbots streamline customer support and engagement, malicious actors exploit similar technology for deceptive purposes. These AI-driven systems mimic natural language, enabling them to convincingly engage with targets and manipulate trust over time. Unlike traditional phishing attempts, which often rely on static scripts and predictable patterns, chatbot-based attacks are dynamic, adaptive, and capable of sustaining long-term conversations, significantly increasing the probability of successful exploitation (Ariza et al., 2023; Falade, 2023).

A critical evolution in this domain is the incorporation of large language models (LLMs) like GPT-based systems, which provide attackers with context-aware responses and the ability to mirror human conversation patterns almost flawlessly. Studies indicate that such attacks are particularly effective on professional networking platforms and dating applications, where users inherently expect conversational interaction (Maurício Ariza et al., 2023). This sophistication makes chatbot attacks not just scalable but also more personalized than ever before, intensifying their social engineering impact.

2.3.4.1 AI-Powered Fake Customer Service or Romance Scams

One of the most concerning manifestations of conversational AI attacks involves impersonating customer service agents or romantic partners. In the fake customer service scenario, attackers deploy chatbots to impersonate representatives of banks, e-commerce platforms, or telecommunication companies. These bots engage victims in realistic dialogues, requesting sensitive details such as login credentials, credit card information, or one-time passwords under the guise of "account verification" or "security checks" (Bharati, 2024; Sabatini, 2025). Unlike traditional email phishing, these interactive exchanges enhance credibility by responding contextually and resolving queries in real time, which significantly lowers suspicion.

Similarly, AI-driven romance scams leverage LLM-powered bots on dating platforms and social media to foster emotional connections with victims over extended periods. These bots demonstrate empathy, maintain consistent personality traits, and even adapt to a victim's communication style. By exploiting psychological vulnerabilities such as loneliness or trust, the attacker can eventually request financial assistance or extract private data for identity theft (Kaur, 2025; Gomathi et al., 2025). Research by Falade (2023) highlights that these romance scams are becoming increasingly automated, reducing the need for direct human intervention while achieving higher success rates.

The adaptive nature of these bots makes detection a formidable challenge. Unlike scripted scams, AI-powered conversational attacks can pivot strategies when faced with suspicion, employ humor or emotional appeals, and even mimic cultural nuances, rendering traditional anti-phishing filters largely ineffective (Schmitt & Flechais, 2023). Consequently, these scams represent a paradigm shift from static deception to dynamic, relationship-based manipulation, which demands advanced defense mechanisms such as human-in-the-loop monitoring and AI-driven anomaly detection frameworks (Gupta et al., 2023).

2.4 Impact on Modern Cybersecurity

The integration of artificial intelligence (AI) into social engineering tactics has significantly altered the cybersecurity threat landscape, resulting in increased attack success rates, accelerated execution speed, psychological exploitation at scale, and substantial financial repercussions. These advancements underscore the growing sophistication of cyberattacks and their potential to undermine traditional security measures.

2.4.1 Increased attack success rate and speed.

AI-driven social engineering attacks exhibit a marked improvement in both success rates and operational speed when compared to conventional techniques. The ability of AI models, such as large language models (LLMs), to analyze vast amounts of personal and organizational data enables attackers to craft highly convincing, contextually relevant messages with minimal effort (Sabatini, 2025; Bharati, 2024). Tools such as FraudGPT and WormGPT demonstrate how AI automates content generation for phishing and spear-phishing campaigns, reducing human error and time traditionally associated with attack preparation (Falade, 2023). Consequently, attackers can execute campaigns on a massive scale while maintaining a high degree of personalization, which historically was a resource-intensive endeavor (Gupta et al., 2023).

2.4.2 Psychological exploitation at scale.

One of the most profound implications of AI-enhanced social engineering lies in its ability to exploit cognitive and emotional biases at an unprecedented scale. Unlike traditional phishing, which often relies on generic messages, AI-driven attacks incorporate behavioral insights and linguistic cues tailored to individual or organizational profiles (Kaur, 2025). These manipulative strategies leverage principles of social psychology such as authority, urgency, and reciprocity to elicit compliance from targets more effectively (Rohini et al., 2025). Furthermore, AI-powered voice cloning and deepfake technologies have expanded attack vectors beyond textual communication, enabling attackers to impersonate trusted individuals through audio and video with remarkable realism (Almutairi & Elgibreen, 2022). This multifaceted approach amplifies the psychological pressure on victims, thereby increasing the likelihood of successful compromise (Schmitt & Flechais, 2023).

2.4.3 Cost of breaches linked to AI-driven attacks.

The financial impact of AI-enhanced social engineering attacks is substantial, as organizations face higher costs associated with detection, remediation, and reputational damage. According to Bharati (2024), the automation capabilities of AI allow attackers to scale their operations with minimal expenditure, creating an asymmetric threat environment where defenders bear disproportionately higher costs to prevent and mitigate breaches. Recent case analyses reveal that AI-generated phishing campaigns significantly reduce the time-to-compromise, leading to faster infiltration of sensitive systems and greater data exfiltration volumes (Falade, 2023). These breaches often necessitate extensive incident response measures, legal compliance efforts, and customer trust restoration, collectively inflating the total cost of cyber incidents (Gupta et al., 2023). In addition, the use of AI in crafting multilingual and culturally adaptive attacks has extended the geographic scope of threats, imposing global economic implications (Kumar & Patel, 2025).

2.5 Defense Mechanisms

As AI-enhanced social engineering attacks become increasingly sophisticated, organizations and individuals face growing challenges in detecting and mitigating these threats. Defensive strategies must evolve to address not only the technical aspects of AI-driven attacks but also the human vulnerabilities they exploit. This section reviews current approaches, including AI-powered detection systems, human-in-the-loop frameworks, awareness and training programs, and policy or regulatory measures. By examining both technical and human-centered defenses, this review highlights strengths, limitations, and opportunities for enhancing resilience against AI-enabled social engineering, particularly in institutional settings such as the University of Benin.

2.5.1 AI for detection vs. AI for attack (arms race).

The cybersecurity landscape has evolved into an adversarial dynamic where artificial intelligence (AI) is utilized both as a defensive and offensive tool. Defensive systems leverage AI-driven threat detection models that employ machine learning to identify phishing patterns, malicious URLs, and anomalous behaviors (Gupta et al., 2023). These systems excel at analyzing vast

datasets, identifying correlations, and reducing detection latency. However, attackers have responded by utilizing generative AI systems such as WormGPT and FraudGPT to craft highly convincing and adaptive phishing campaigns (Falade, 2023; Sabatini, 2025). This has resulted in a technological arms race where improvements in defensive algorithms are rapidly counteracted by more sophisticated attack strategies.

Unlike traditional phishing, which often relied on grammatical errors and generic templates, AI-powered attacks generate content indistinguishable from legitimate communications (Schmitt & Flechais, 2023). Furthermore, AI enables real-time adaptation, allowing attackers to modify messages based on initial victim responses, thus bypassing static filtering mechanisms (Bharati, 2024). Consequently, while AI has enhanced defense capabilities, it has also created an environment of perpetual escalation between attackers and defenders.

2.5.2 Limitations of current phishing filters.

Despite advancements in phishing detection, current filtering technologies face significant limitations. Conventional email security systems rely heavily on signature-based detection, blocklists, and heuristics to identify malicious content (Kaur, 2025). However, generative AI circumvents these measures by producing contextually relevant and unique content that does not match pre-existing signatures (Patel & Kumar, 2025). Moreover, attackers can dynamically alter URLs and hosting environments to avoid blacklisting, further complicating detection efforts.

AI-driven filters that rely on natural language processing (NLP) and anomaly detection are also vulnerable to adversarial manipulation. For example, small perturbations in syntax or the strategic inclusion of benign language elements can reduce the likelihood of detection by AI models (Bharati, 2024). Furthermore, while deep learning-based filters offer higher detection accuracy, they are resource-intensive and may introduce latency issues, limiting real-time scalability in enterprise environments (Sabatini, 2025).

2.5.3 Human factor training vs. AI sophistication.

Organizations have historically emphasized user awareness and training as a primary defense against phishing attacks. Training programs aim to educate employees on recognizing suspicious

emails, avoiding link-clicking, and verifying sender identities (Rohini et al., 2025). While such measures remain crucial, their effectiveness diminishes against AI-powered attacks that exploit psychological and cognitive biases with unprecedented precision (Falade, 2023). Generative AI systems can craft personalized messages based on publicly available data, making them appear highly authentic and reducing the likelihood of user skepticism (Schmitt & Flechais, 2023).

Moreover, the scalability of AI-driven phishing attacks exacerbates the challenge. Attackers can generate millions of unique, context-aware emails at minimal cost, overwhelming traditional human-centric defense strategies (Bharati, 2024). This imbalance highlights the insufficiency of relying solely on human vigilance in an era where attackers employ adaptive, machine-driven techniques. Consequently, organizations must transition from reactive awareness programs to proactive AI-augmented defenses that integrate predictive analytics and real-time threat intelligence (Gupta et al., 2023).

2.6 Summaries of past studies and related works

To consolidate existing empirical evidence on AI-enhanced social engineering attacks, it is important to examine studies that have investigated related phenomena in various contexts. While individual studies provide valuable insights, a comparative overview allows for clearer identification of trends, methodologies, and findings, as well as their relevance to the current study. The table below presents a synthesis of selected studies, highlighting their focus, key findings, and applicability to understanding AI-driven social engineering attacks within university environments.

Table 1: A comparative summary of relevant literature

S/N	Author(s)	Year	Study Focus	Key Findings	Relevance to This Study
1	Rahul Kailas Bharati	2024	AI-Enhanced Social Engineering: Evolving Tactics in Cyber Fraud and Manipulation	Identified that AI tools enable large-scale personalized phishing and impersonation schemes through automation and deep learning.	Provides foundational insight into how AI has transformed traditional social engineering into more sophisticated attacks.

S/N	Author(s)	Year	Study Focus	Key Findings	Relevance to This Study
2	T. F. Blauth, O. Gstrein & A. Zwitter	2022	Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI	Highlighted the misuse of AI in cybercrime, including automated scams, misinformation, and privacy violations.	Supports the argument that AI is a dual-use technology with both beneficial and malicious potential in cybersecurity contexts.
3	Polra Victor Falade	2023	Decoding the Threat Landscape: ChatGPT, FraudGPT, and WormGPT in Social Engineering Attacks	Demonstrated that generative AI tools are now weaponized to automate phishing and impersonation, reducing attacker effort and increasing success rates.	Directly informs this study's focus on generative AI tools' involvement in real-world social engineering.
4	Brando Marzio Sabatini	2025	The Next Generation of Cyber Threats: AI in Social Engineering Attacks	Explained that the integration of AI enhances deception and adaptive manipulation in phishing and voice cloning attacks.	Strengthens the theoretical background on how AI deepens the sophistication of human-targeted cyber threats.
5	Dr. S. Gomathi Alias Rohini et al.	2025	Social Engineering Attacks in Cybersecurity	Classified types of social engineering techniques and prevention measures, emphasizing human vulnerability.	Provides a conceptual basis for understanding attack types and human susceptibility within the UNIBEN context.
6	Bhupinder Kaur	2025	Social Engineering Attacks in the Digital Age	Found that increasing digital dependence and lack of awareness amplify vulnerability to online scams.	Reinforces the study's finding that despite awareness, exposure and victimization rates remain high among digital natives.
7	Nishant Kumar & Niyati Manojkumar Patel	2025	Social Engineering Attack in the Era of Generative AI	Analyzed the impact of generative AI in facilitating identity-based fraud and deepfake manipulation.	Directly supports the section on AI's role in deception through synthetic content and impersonation.
8	Marc Schmitt & Ivan Flechais	2023	Digital Deception: Generative Artificial Intelligence in Social Engineering and Phishing	Reported that generative AI models can create hyper-realistic phishing content indistinguishable from legitimate communication.	Strengthens the argument that detection and awareness alone are insufficient against AI-powered deception.

S/N	Author(s)	Year	Study Focus	Key Findings	Relevance to This Study
9	Zaynab Almutairi & Hebah Elgibreen	2022	A Review of Modern Audio Deepfake Detection Methods: Challenges and Future Directions	Identified challenges in detecting AI-generated audio used in fraud and impersonation attacks.	Supports the recommendation for AI-integrated defense systems that can detect deepfake and voice-based scams.
10	Maurício Ariza et al.	2023	Automated Social Engineering Attacks using ChatBots on Professional Social Networks	Showed that AI chatbots can autonomously conduct targeted social engineering attacks with contextual accuracy.	Reinforces the study's premise that automation increases attack frequency and sophistication.
11	Maanak Gupta et al.	2023	From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy	Discussed both the benefits and emerging threats of generative AI, highlighting the privacy and security implications.	Provides the broader cybersecurity framework within which AI-enhanced threats should be analyzed.
12	Human-in-the-Loop AI Framework (Author unspecified)	—	Human-in-the-Loop AI for Defending Against Social Engineering Attacks	Proposed a hybrid model integrating human judgment with AI tools for adaptive defense.	Inspired the Human–AI Cyber Awareness Framework (HACAF) proposed in this research as a mitigation model.

From this comparative review, it is evident that while significant research has been conducted on AI-enhanced social engineering attacks globally, there is limited focus on higher education environments in developing countries, particularly in Nigeria. Furthermore, existing studies often emphasize either offensive techniques or defense mechanisms, but rarely both in a single context. This gap highlights the importance of the present study, which seeks to explore both the awareness and susceptibility of students to AI-powered attacks at the University of Benin while proposing a Human–AI Cyber Awareness Framework (HACAF) to strengthen institutional resilience against such threats.

2.7 Future Trends and Research Gaps

The landscape of AI-enhanced social engineering is rapidly evolving, driven by continuous advancements in artificial intelligence and machine learning. Future trends indicate that attackers will increasingly exploit these technologies to create more adaptive, scalable, and context-aware attacks. Several emerging patterns and developments are worth highlighting:

1. **Integration of Multimodal AI in Attacks**: Future social engineering campaigns are expected to leverage multimodal models that combine text, audio, video, and image synthesis. Deepfake-driven impersonations may no longer be limited to voice cloning but will include fully animated avatars capable of real-time interaction, complicating detection (Almutairi & Elgibreen, 2022).

2. **AI-Augmented Social Bots with Emotional Intelligence**: Advanced conversational models are anticipated to include affective computing capabilities, enabling bots to detect and exploit emotional states during interactions, thereby improving persuasion and manipulation success rates (Bharati, 2024).

3. **Personalization at Scale through Federated Learning**: Attackers may employ federated learning and on-device models to mine sensitive behavioral patterns without centralized data collection, enabling highly individualized phishing or scam messages while evading conventional detection methods (Kumar & Patel, 2025).

4. **Weaponization of Autonomous AI Agents**: The rise of autonomous AI agents capable of planning, executing, and adapting attacks without human intervention poses a significant threat. These agents could self-improve and dynamically adjust strategies based on victim response, creating persistent attack ecosystems (Gupta et al., 2023).

Despite these advancements, significant research gaps remain:

2.7.1 Research Gaps

1. Most existing studies examine traditional social engineering attacks or AI capabilities separately, leaving limited understanding of the combined effect of AI-enhanced attacks on human behavior and organizational systems (Falade, 2023; Bharati, 2024).
2. There is a lack of research on real-world user experiences with AI-driven attacks, particularly regarding detection challenges, psychological impact, and contextual effectiveness (Kumar & Patel, 2025).
3. Many studies rely on generalized or global data, with minimal attention to specific institutional or regional contexts, especially in developing countries such as Nigeria.
4. There is a need for practical frameworks that integrate human cognition with AI-based defenses to mitigate sophisticated AI-driven social engineering attacks (Gupta et al., 2023).

2.7.2 How This Research Differs

This research addresses these gaps by focusing on AI-enhanced social engineering attacks with a multi-dimensional analysis, integrating technical, behavioral, and cognitive aspects. Unlike prior studies that primarily discuss detection techniques or describe attacks in isolation, this work:

1. Focuses specifically on AI-enhanced social engineering attacks within the University of Benin, providing a localized and institutional perspective.
2. Combines theoretical analysis, empirical survey data, and contemporary case studies to examine both human and technical vulnerabilities in a single study.
3. Unlike prior research, it emphasizes practical insights from users' experiences with AI-driven threats across multiple channels, including email, social media, and conversational AI.
4. The study aims to develop context-specific recommendations and frameworks to improve cybersecurity awareness and resilience within institutional settings, complementing global research findings.

5. By integrating human factors with technical analyses, this study addresses the dual challenge of detection and response, which is often overlooked in existing literature.

2.8 Summary

This chapter reviewed existing literature on social engineering and the evolving role of artificial intelligence in cyberattacks. It highlighted how traditional social engineering exploits human psychology and trust to bypass security systems and how AI, particularly generative models and machine learning, has amplified the scale, sophistication, and personalization of these attacks. Key techniques such as AI-driven phishing, deepfake-based impersonation, voice cloning, and chatbot-based attacks were examined, demonstrating their capacity to manipulate human behavior and evade conventional defenses.

The chapter also explored the evolution of AI in cyberattacks, from early automation to modern generative models, emphasizing the increasing complexity and adaptive nature of these threats. Emerging trends, such as multimodal attacks, AI-augmented social bots, and autonomous AI agents, were discussed to show the growing challenge for defenders.

Research gaps identified include the limited understanding of user experiences with AI-enhanced attacks, underexplored cross-modal threats, and the lack of context-specific frameworks for detection and mitigation, particularly in institutions in developing countries. This study differentiates itself by focusing on the University of Benin, integrating theoretical, empirical, and practical perspectives, and aiming to develop actionable recommendations to strengthen cybersecurity awareness and resilience.

In conclusion, the literature underscores that while AI has significantly advanced offensive capabilities, there is a pressing need for research and strategies that integrate human and technical defenses to effectively counter AI-enhanced social engineering threats.

CHAPTER 3

RESEARCH METHODOLOGY

3.0 Introduction

Research methodology is a systematic plan that guides how a study is conducted to ensure that its findings are credible, valid, and relevant. This chapter explains the methods and procedures adopted to achieve the objectives of this study titled "An Analytical Study of AI-Enhanced Social Engineering Attacks and Their Impact on Modern Cybersecurity: A Case Study of the University of Benin (UNIBEN), Nigeria."

The chapter describes how data were collected, organized, and analyzed to answer the research questions and test the research objectives. It also outlines the type of research design used, the study population, the sampling methods, and the instruments applied in data gathering. Since this research investigates human behavior in response to AI-driven cyber threats, it employs a combination of both quantitative and qualitative methods.

The use of mixed methods is essential because AI-enhanced social engineering is not only a technical issue but also a psychological and social one. Quantitative methods help measure the level of awareness, exposure, and preparedness among respondents, while qualitative methods help to explain why and how these attacks are effective. This methodological combination provides a balanced understanding of both the technical and human aspects of the problem.

Additionally, this chapter highlights the steps taken to ensure ethical compliance and the credibility of data collected from participants within the University of Benin. The university setting was selected as the case study because it represents a typical academic environment where digital systems, online communication, and human interaction are deeply intertwined factors that make it an ideal environment to explore AI-driven social engineering threats.

By the end of this chapter, the reader will clearly understand how this study was conducted, the reasoning behind each methodological choice, and how the process ensures that the results can be trusted and replicated in future studies.

3.1 Research Design

The research design provides the structural framework and strategy that guided this study. It determines how data were collected, analyzed, and interpreted to address the research objectives effectively. For this study, a descriptive and analytical survey design was employed.

A descriptive survey design was considered appropriate because it enables the researcher to systematically gather information from a large population to describe current conditions and opinions. This design helps in identifying patterns and relationships among variables without manipulating them (Creswell, 2014). In this study, it was used to capture participants' awareness, perception, and vulnerability to AI-enhanced social engineering attacks within the University of Benin community.

The analytical component complements this by allowing the researcher to interpret relationships between variables such as digital literacy, cybersecurity awareness, and exposure to AI-driven manipulation (Saunders et al., 2019). This design, therefore, fits well with the study's dual aim of description and explanation showing not only what is happening but also why certain cybersecurity behaviors persist.

This approach is particularly relevant given the complex nature of AI-enhanced social engineering, where psychological, technological, and organizational factors intersect (Bharati, 2024; Falade, 2023). It provides a systematic method for linking theoretical insights from existing literature to practical realities observed within a university environment.

The survey method also allows the use of structured questionnaires, which are effective for collecting quantifiable and comparable data from a large sample of respondents (Kumar, 2019). With a sample size of 140 participants, the study ensures a sufficient representation of both staff and students, thereby strengthening the reliability and generalizability of findings.

In summary, this research design was selected because it supports both descriptive and inferential analyses, ensuring that the results are evidence-based and reflective of how AI-driven social engineering affects cybersecurity awareness and response in academic institutions.

3.2 Population of the Study

The population of a study refers to the entire group of individuals that share similar characteristics relevant to the research and from which the sample is drawn (Kumar, 2019). For this research, the population comprises undergraduate students of the University of Benin (UNIBEN), Nigeria). Students were chosen because they are among the most active users of digital platforms, including online learning systems, social media, and email, all of which are frequent targets of social engineering and phishing attacks.

The University of Benin is one of Nigeria's leading federal universities, with an estimated student population of about 60,000 across its various faculties and departments (University of Benin Official Website, 2024). Students at UNIBEN frequently engage in online academic registration, e-learning platforms, and digital communications, which exposes them to both legitimate and malicious uses of artificial intelligence technologies. This makes them a suitable demographic for studying how AI-enhanced social engineering attacks exploit human trust and digital behavior (Bharati, 2024; Falade, 2023).

The study population was drawn from four key faculties that represent a balance between technical and non-technical disciplines, ensuring diversity in digital exposure and awareness levels:

1. Faculty of Physical Sciences (including Computer Science, Mathematics, and Statistics): where students typically possess stronger digital literacy and a foundational understanding of AI systems.
2. Faculty of Social Sciences: whose students frequently use digital tools for collaboration and research, making them susceptible to phishing and information manipulation.
3. Faculty of Engineering: where students often interact with technical infrastructures and emerging technologies, exposing them to practical cybersecurity concerns.
4. Faculty of Management Sciences: comprising students who regularly handle finance-related coursework and transactions, which makes them attractive targets for AI-assisted financial scams and fraudulent schemes.

Students from 100 to 500 levels were included to capture a range of experiences, knowledge, and digital literacy levels. Lower-level students (100–200) may be less informed about cybersecurity risks, while higher-level students (300–500) may have more awareness through exposure to research and professional training. This stratified inclusion allows for a deeper analysis of how academic level and field of study influence vulnerability to AI-driven manipulation (Rohini et al., 2025).

The decision to focus on students rather than staff is based on the observation that young adults are both early adopters of digital tools and prime targets for online deception due to trust bias and overconfidence (Gupta et al., 2023). Consequently, this population provides valuable insights into the intersection between AI-generated deception techniques and human susceptibility within an academic digital environment.

3.3 Sample and Sampling Technique

A sample represents a smaller group selected from the population to participate in a study, allowing the researcher to make generalizations about the entire population (Kothari, 2014). In this study, the sample was drawn from the undergraduate student population of the University of Benin (UNIBEN), estimated at 60,000 students across all faculties (University of Benin, 2024). Due to time and resource limitations, it was not feasible to study the entire population. Therefore, a representative sample was carefully selected to reflect the diversity of students in terms of faculty, department, and level of study.

The sample size for this study was determined using the Yamane (1967) formula, which is suitable for finite populations and helps ensure a statistically acceptable level of accuracy. The formula is expressed as:

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

Where:

- n = sample size
- N = population size (60,000)
- e = margin of error (0.08 or 8%)

Substituting the values:

$$n = \frac{60,000}{1 + 60,000(0.08)^2} = \frac{60,000}{1 + 384} \approx 155$$

Figure 1: Illustration of Sample Size Determination Using the Yamane Formula

Due to time and logistical constraints, a final sample of 140 valid responses was analyzed, which remains statistically acceptable for descriptive social research (Kumar, 2019).

3.3.1 Sampling Technique

The stratified random sampling technique was adopted to ensure fair representation of students from different academic levels and faculties. The population was first divided into strata based on faculty including the Faculty of Physical Sciences, Faculty of Social Sciences, Faculty of Engineering, and Faculty of Management Sciences since these faculties reflect both technical and non-technical backgrounds. Within each stratum, students were randomly selected to participate in the survey.

This method minimizes sampling bias and ensures that the views captured reflect the varying degrees of technological exposure and cybersecurity awareness among students (Creswell & Creswell, 2018). The inclusion of students from both lower and upper levels (100–500 level) also allowed the study to explore how exposure and experience affect vulnerability to AI-enhanced social engineering attacks.

Furthermore, participation was voluntary, and respondents were selected based on their willingness to complete the online questionnaire. The link to the survey was distributed through official university social platforms and student WhatsApp groups. This approach was chosen because it ensured accessibility, convenience, and broad participation within the university community (Blauth et al., 2022).

Overall, the use of stratified random sampling combined with the Yamane formula enhanced the reliability and representativeness of the data collected, providing a sound basis for valid conclusions.

3.4 Method of Data Collection

Data for this study were collected using a structured online questionnaire designed to obtain both qualitative and quantitative information from undergraduate students of the University of Benin. The questionnaire served as the main instrument for gathering primary data, while relevant textbooks, journals, and online publications were reviewed to provide secondary data.

The online questionnaire was developed using Google Forms and distributed digitally through WhatsApp and Telegram groups commonly used by students. This method was selected because it allowed easy access to respondents across different faculties, reduced printing costs, and provided automatic recording of responses, which minimized errors in data entry (Creswell & Creswell, 2018; Saunders et al., 2019).

The questionnaire was divided into five major sections (A–E), each aligned with the research objectives and questions of the study. These sections are summarized below:

Section A: Basic Information

This section gathered background data on respondents such as age, gender, occupation, and internet usage frequency. These demographic details helped the researcher to understand the diversity and online exposure level of participants, which may influence their vulnerability to social engineering attacks.

Section B: Awareness of AI and Online Scams

This part assessed the respondents' awareness of social engineering and AI-generated scams

such as fake chatbots, deepfake videos, or phishing emails. It also explored the platforms where respondents mostly encounter suspicious activities, helping to measure the spread and visibility of AI-driven deception in students' daily online environments.

Section C: Experience and Exposure

This section examined whether respondents had ever been victims of hacking, scams, or other deceptive activities online. Follow-up questions asked about the nature of these incidents (e.g., through email, fake websites, or voice calls) and whether the scams appeared to be AI-generated. This data was useful in understanding the direct impact and personal experiences students have had with AI-enhanced social engineering attacks.

Section D: Knowledge and Safety Practices

This section tested the respondents' knowledge of cybersecurity practices and their preventive measures, such as using strong passwords, two-factor authentication, and avoiding unknown links. It also assessed whether students had ever attended any form of cybersecurity or AI safety awareness training. These responses helped measure how prepared students are to defend themselves against modern AI-based scams.

Section E: Opinions and Recommendations

The final section collected respondents' opinions about the seriousness of AI-enhanced social engineering threats, their confidence in identifying AI-generated scams, and their views on how effective current security systems are. The section ended with an open-ended question asking for respondents' suggestions on how to protect people from AI-powered online scams. This qualitative input provided useful insights and personal perspectives to complement the numerical data collected.

To ensure accuracy and reliability, the questionnaire was pretested with 10 students before final distribution. Their feedback was used to refine the wording of questions, correct ambiguity, and ensure that the items were simple and easy to understand. After this pilot test, the final questionnaire was distributed, resulting in 140 valid responses that were analyzed for this study.

All participants were informed about the purpose of the research, assured of the confidentiality of their responses, and notified that participation was voluntary. No identifying information was

collected, ensuring the protection of participants' privacy and compliance with ethical standards (Saunders et al., 2019).

In summary, the questionnaire design and administration ensured that the data collected were reliable, comprehensive, and directly relevant to understanding students' awareness, experience, and preparedness against AI-enhanced social engineering attacks at the University of Benin.

3.5 Method of Data Analysis

The data collected from the administered questionnaires were carefully reviewed, coded, and analyzed to address the research objectives and questions. Since the study employed a descriptive survey design, the analysis focused mainly on identifying trends, patterns, and relationships within the responses. The responses obtained from Google Forms were automatically exported to Microsoft Excel (version 2021) for organization and statistical processing.

In Excel, responses were categorized and coded according to the sections of the questionnaire covering , awareness, experience, safety practices, and opinions regarding AI-enhanced social engineering attacks. Descriptive statistical tools such as frequency counts, percentages, and mean scores were employed to summarize the data, while graphical representations such as bar charts and pie charts were used to illustrate key findings.

This approach enabled a clear and straightforward interpretation of the data, allowing the researcher to present results in an accessible and visually comprehensible format. The use of Microsoft Excel was considered adequate due to its efficiency in handling survey-based datasets, ease of visualization, and reliability for descriptive analysis.

The analyzed data formed the basis for interpreting patterns related to the level of awareness, exposure, and perception of AI-enhanced social engineering attacks among students of the University of Benin. Findings were later compared with insights drawn from existing literature to establish areas of agreement or deviation.

3.6 Characteristics of the Study

This study focused on examining how artificial intelligence (AI) is influencing the rise and complexity of social engineering attacks, using students of the University of Benin as the case population. The research was characterized by its emphasis on human behavior, awareness levels, and interaction with AI-driven online threats. It adopted a quantitative and descriptive research design, which is considered effective for analyzing behavioral patterns and perceptions within a defined group (Kaur, 2025; Gomathi Alias Rohini et al., 2025).

One major characteristic of this study is its cross-sectional nature, meaning that data were collected from respondents at a single point in time rather than over an extended period. This approach is commonly used in cybersecurity behavior studies to provide a snapshot of awareness and exposure levels among a specific population (Bharati, 2024).

Another defining feature of the study is its use of structured questionnaires as the main data collection instrument. The questionnaire, divided into five sections, captured demographic information, awareness of AI scams, exposure experiences, safety practices, and personal opinions. This design aligns with previous research emphasizing the role of structured surveys in evaluating cybersecurity and AI awareness (Falade, 2023; Ariza et al., 2023).

The study also follows a descriptive analytical framework, focusing on summarizing observed patterns and tendencies rather than testing complex hypotheses. Such frameworks are particularly suitable for identifying trends in emerging areas like AI-enhanced social engineering, where behavioral responses and awareness levels are still developing (Gupta et al., 2023; Schmitt & Flechais, 2023).

Furthermore, the research maintained objectivity and ethical neutrality by treating all respondents equally and analyzing results using verifiable, quantitative methods. This ensures reliability and transparency qualities that are essential in cybersecurity-related human studies (Sabatini, 2025).

In summary, this study is characterized by its focus on AI-enhanced social engineering, its student-centered sample, and its use of descriptive and ethical research methods. Together, these elements enable the research to contribute meaningfully to the growing body of work exploring

how AI-driven manipulation influences online safety and human vulnerability in modern digital environments.

3.7 Advantages of the Study

This research presents several advantages that contribute both to academic scholarship and to practical cybersecurity awareness within higher institutions such as the University of Benin. One of the primary strengths lies in its focus on AI-enhanced social engineering attacks, a topic that remains underexplored in the Nigerian academic context despite its growing global relevance (Bharati, 2024; Sabatini, 2025). By centering the investigation on university students a population that represents active internet users and a major target group for online scams the study captures the perspectives of individuals who are most exposed to AI-driven social manipulation. This demographic relevance increases the reliability of the findings in portraying real-world awareness levels and behavioral tendencies (Kaur, 2025).

Another key advantage is the integration of both traditional and emerging cyber threat dimensions within a single analytical framework. While previous works have largely treated social engineering as a conventional psychological manipulation problem (Rohini et al., 2025), this study expands the lens to include the impact of generative AI tools such as ChatGPT, FraudGPT, and WormGPT (Falade, 2023). This dual approach allows for a more comprehensive understanding of how artificial intelligence amplifies deception, personalization, and automation in social engineering, filling a significant conceptual gap identified in earlier research.

Methodologically, the study's use of a structured questionnaire designed around awareness, exposure, and mitigation behaviors provides a clear and quantifiable representation of the cybersecurity attitudes of university students. The instrument's design draws inspiration from empirical surveys in similar studies (Ariza et al., 2023; Gupta et al., 2023), enhancing its credibility and comparability. The adoption of Google Forms for online administration ensured accessibility, convenience, and anonymity factors that encouraged participation and reduced response bias.

Furthermore, the application of descriptive and inferential statistical analysis using modern analytical tools strengthens the study's validity. Unlike narrative or purely conceptual reviews, this research incorporates data-driven insights that can be used to inform institutional cybersecurity policy and awareness programs (Schmitt & Flechais, 2023). The quantitative emphasis also complements existing qualitative literature, offering measurable patterns of AI-related scam exposure and digital safety practices.

Lastly, the study provides practical value beyond academic contribution. Its findings can guide policy formulation, campus cybersecurity training, and digital literacy initiatives within the University of Benin and other Nigerian universities. By revealing the extent of awareness and the gaps in protective behavior, the research contributes to the design of AI-resilient security awareness models a growing necessity in an age where artificial intelligence continually reshapes the threat landscape (Kumar & Patel, 2025; Bharati, 2024).

3.8 Ethical Considerations

Ethical responsibility was a major priority throughout this research, especially since it involved human participants and sensitive discussions about online scams and cyberattacks. The study strictly followed standard ethical guidelines for social science research to ensure honesty, respect, and protection of participants' privacy and data (Blauth et al., 2022).

To begin with, informed consent was obtained from all participants before they took part in the survey. Each respondent was given a short explanation of the study's purpose, its voluntary nature, and their right to withdraw at any time without any consequence. This ensured that participation was based on full understanding and agreement, in line with international research ethics principles such as those outlined in the Belmont Report and by the American Psychological Association (APA, 2022).

Another ethical concern was confidentiality. Since the topic of online scams could make some respondents feel embarrassed or exposed, all responses were kept strictly anonymous. The questionnaire did not collect names, student identification numbers, or any information that could trace responses back to individuals. Data was stored securely on Google Forms and later

downloaded to a password-protected file accessible only to the researcher. This procedure reduced the risk of data misuse or unauthorized access (Kaur, 2025).

The research also respected non-maleficence, meaning no harm should come to any participant. Questions were carefully phrased to avoid psychological distress, judgment, or intimidation. Respondents were not required to disclose personal experiences that made them uncomfortable. Additionally, since the topic involves AI-generated deception, the study avoided any form of simulated or deceptive activity that could mislead participants, maintaining full transparency throughout the data collection process (Gupta et al., 2023).

In line with academic integrity, plagiarism was avoided by properly acknowledging all authors and sources used in the review of literature and theoretical discussions. Every secondary material was paraphrased in simple and original language, and all sources were cited according to recognized referencing standards (Falade, 2023; Sabatini, 2025).

Finally, ethical clearance was conceptually guided by the University of Benin's code of research conduct, which emphasizes the dignity, safety, and consent of participants in social research. The study's design and online administration were tailored to fit this framework, ensuring compliance with institutional and national standards for ethical research (Rohini et al., 2025).

In summary, all ethical measures consent, confidentiality, data protection, transparency, and respect for participants were rigorously upheld to maintain trust, credibility, and academic integrity.

3.9 Tools Used

The primary tool used for data analysis in this study was Microsoft Excel 2021. After data collection through Google Forms, the responses were automatically exported to Excel for cleaning, organization, and analysis. The software facilitated the computation of frequency distributions, percentages, and the creation of charts and tables that summarize respondents' awareness, experiences, and perceptions of AI-enhanced social engineering attacks.

Excel was also used to perform cross-tabulations where necessary, allowing for comparisons across demographic variables such as age, gender, and internet usage frequency. The choice of

Excel was based on its accessibility, simplicity, and versatility in handling quantitative data without requiring advanced programming or statistical expertise. According to Prajapati (2020) and Kumar & Mishra (2023), Microsoft Excel remains one of the most efficient tools for descriptive statistical analysis in small and medium-scale research projects, especially when the focus is on summarizing and visualizing trends rather than performing complex inferential tests.

By using Excel, the researcher ensured that the analysis process remained transparent, replicable, and aligned with the study's descriptive analytical design. The results obtained were subsequently presented in tabular and graphical formats in Chapter Four to aid clear interpretation and discussion.

CHAPTER 4

ANALYSIS AND RESULTS

4.0 Introduction

This chapter presents the analysis and interpretation of data collected from 140 respondents at the University of Benin (UNIBEN). The results are organized in line with the research objectives, which focused on understanding the awareness, experience, and perception of AI-enhanced social engineering attacks among students.

Both descriptive and inferential statistical methods were used to analyze the responses. Descriptive analysis summarizes the data using frequency tables, percentages, and charts, while inferential analysis explores the relationship between selected variables. The findings are then discussed in relation to the reviewed literature to highlight key insights and implications for cybersecurity awareness and education.

4.1 Presentation of Data

This section presents the results obtained from the survey conducted among students of the University of Benin (UNIBEN). A total of 140 respondents participated, providing insights into their demographics, social media usage, awareness of social engineering attacks, experience with scams, and perceptions of AI-enhanced attacks. The data is presented in tables and charts to facilitate understanding.

4.1.1 Age Distribution of Respondents

The age distribution of participants is shown in Table 4.1. The majority of respondents (90%) are between 18 and 24 years, reflecting the predominant age group of university students. Respondents aged 25–34 make up 9.3%, while only one respondent (0.7%) is above 55 years. No respondents were recorded in the 35–44 and 45–54 age brackets.

Table 2: Age Distribution of Respondents

Age Group	**Frequency**	**Percentage**
18–24	126	90%
25–34	13	9.3%
35–44	0	0%
45–54	0	0%
55+	1	0.7%
Total	140	100%

4.1.2 Gender Distribution of Respondents

Table 4.2 shows that 52.1% of respondents are male, while 47.9% are female, indicating a fairly balanced gender representation.

Table 3: Gender Distribution of Respondents

Gender	**Frequency**	**Percentage**
Male	73	52.1%
Female	67	47.9%
Total	140	100%

4.1.3 Frequency of Internet/Social Media Use

Respondents were asked how often they use the internet or social media. Table 4.3 shows that almost all respondents (99.3%) use social media or the internet daily, while only one respondent (0.7%) uses it a few times a week. No respondents reported rare or zero usage.

Table 4: Frequency of Internet/Social Media Use

Frequency of Use	**Frequency**	**Percentage**
Every day	139	99.3%
A few times a week	1	0.7%
Rarely	0	0%
Never	0	0%
Total	140	100%

4.1.4 Awareness of Social Media Scams

Table 4.4 shows that the majority of respondents (95%) are aware of social media scams, while 5% are not.

Table 5: Awareness of Social Media Scams

Response	**Frequency**	**Percentage**
Yes	133	95%
No	7	5%

Total	140	100%

4.1.5 Types of Scams Encountered

Respondents were asked about the types of scams they have seen or experienced online. Table 4.5 indicates that fake social media accounts (90%) and WhatsApp/Telegram scams (87.1%) are the most commonly encountered. Phishing emails were experienced by 68.6%, deepfake videos or voice calls by 65.7%, and job or investment scams by 82.9%.

Table 6: Types of Scams Encountered

Type of Scam	**Frequency**	**Percentage**
Phishing emails	96	68.6%
Fake social media accounts	126	90%
WhatsApp/Telegram scams	122	87.1%
Job or investment scams	116	82.9%
Deepfake videos or voice calls	92	65.7%
Other	5	3.5%

4.1.6 Prior Victimization of Social Engineering

Table 4.6 shows that 56.4% of respondents have previously fallen victim to social engineering attacks, while 38.6% have not, and 5% are not sure.

Table 7: Prior Victimization of Social Engineering

Response	**Frequency**	**Percentage**
Yes	79	56.4%
No	54	38.6%
Not sure	5	5%
Total	140	100%

4.1.7 Perception of AI Involvement in Scams

Respondents were asked how easy they find it to differentiate between AI-generated scams and human-generated scams. Table 4.7 indicates that 41.4% find it somewhat easy, 28.6% find it hard, 10.7% find it very easy, 5.7% find it very hard, and 13.6% are not sure.

Table 8: Perception of AI Involvement in Scams

Response	**Frequency**	**Percentage**
Very easy	15	10.7%
Somewhat easy	58	41.4%
Hard	40	28.6%
Very hard	8	5.7%
Not sure	19	13.6%
Total	140	100%

4.2 Descriptive Analysis

The descriptive analysis provides an interpretation of the survey data presented in Section 4.1. It focuses on trends, patterns, and observations regarding respondents' demographics, social media usage, awareness, experience with scams, and perception of AI involvement.

4.2.1 Age Distribution

The majority of respondents (90%) are between 18 and 24 years, which aligns with the predominant age group of undergraduate students at UNIBEN. Only 9.3% are aged 25–34, and a single respondent (0.7%) is above 55 years, while no respondents fall in the 35–44 or 45–54 age brackets. This indicates that the survey primarily captures the perspective of young adults, who are typically the most active on social media and internet platforms.

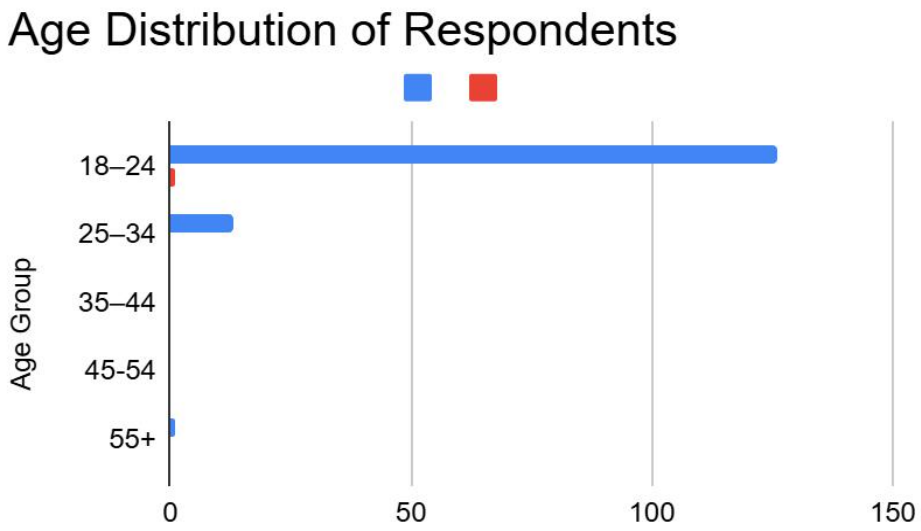

Figure 1: Bar chart representing the age distribution of survey respondents at the University of Benin

4.2.2 Gender Distribution

The gender distribution is fairly balanced, with 52.1% male and 47.9% female respondents. This balance ensures that the results are not heavily skewed toward a particular gender, supporting a representative understanding of awareness and exposure to social engineering attacks among students.

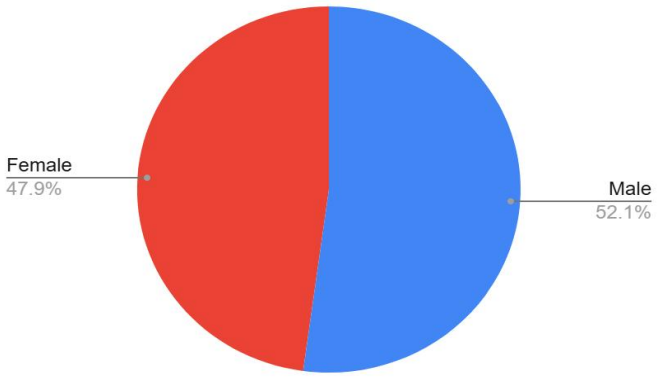

Figure 2: Pie chart representing the gender distribution of survey respondents

4.2.3 Internet and Social Media Use

A very high percentage of respondents (99.3%) report using the internet or social media daily, reflecting the high connectivity of university students. This frequent usage suggests a higher likelihood of exposure to AI-enhanced social engineering attacks and underlines the relevance of studying this population.

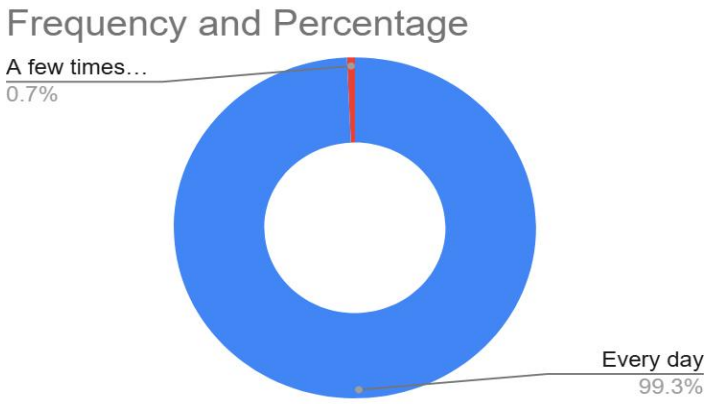

Figure 3: Pie chart representing Frequency of internet and social media use among respondents

4.2.4 Awareness of Social Media Scams

Most respondents (95%) are aware of social media scams, indicating a generally high level of knowledge about online threats. However, 5% of respondents remain unaware, highlighting a small but notable gap in awareness. This aligns with the research objective of understanding the level of exposure and readiness of students to counter AI-driven scams.

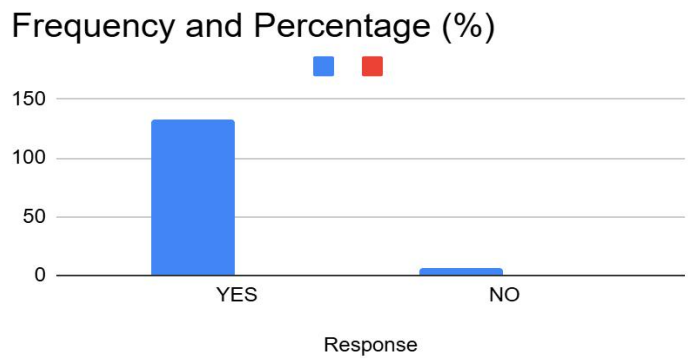

Figure 4: Bar chart representing the awareness of social media scams among respondents

4.2.5 Types of Scams Encountered

Fake social media accounts (90%) and WhatsApp/Telegram scams (87.1%) are the most commonly observed scam types. Phishing emails are experienced by 68.6%, while deepfake videos or voice calls are reported by 65.7% of respondents. Job or investment scams affect 82.9% of respondents. These results suggest that students are encountering a mix of traditional and AI-enhanced scams, particularly on messaging platforms and social networks.

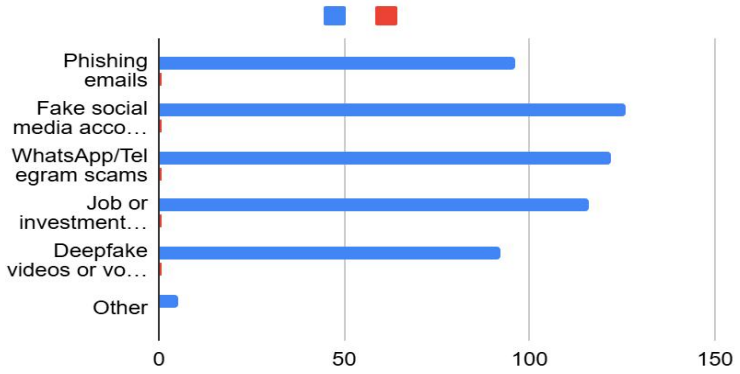

Figure 5: Bar chart representing the types of scams encountered by respondents

4.2.6 Prior Victimization

More than half of the respondents (56.4%) have previously fallen victim to social engineering attacks. This high proportion highlights the persistent vulnerability of students to online deception, reinforcing the importance of raising awareness and improving cybersecurity practices.

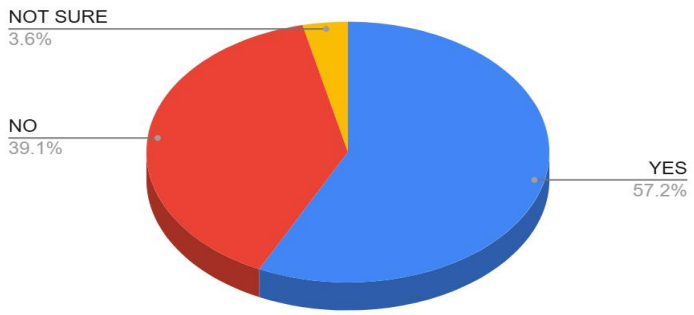

Figure 6: Pie chart representing the prior victimization experience of respondents regarding social engineering attacks

4.2.7 Perception of AI Involvement

Respondents find it somewhat challenging to distinguish AI-generated scams from human-generated ones. While 41.4% find it somewhat easy, 28.6% find it hard, 10.7% find it very easy, 5.7% find it very hard, and 13.6% are unsure. This shows that AI-generated scams are increasingly sophisticated, confirming the research objective to assess the impact of AI on social engineering effectiveness.

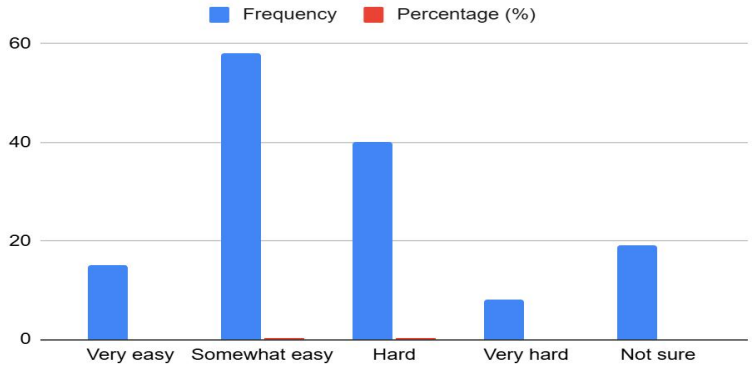

Figure 7: Bar chart representing the respondents' perception of ease in identifying AI-generated scams

Summary of Trends

1. Students aged 18–24 are the most active group and thus the most exposed to online scams.
2. Balanced gender representation ensures reliability of findings across male and female respondents.

3. Daily social media use is nearly universal, implying high susceptibility to AI-powered attacks.
4. Awareness is generally high, yet prior victimization remains significant, indicating that awareness alone does not guarantee protection.
5. AI-enhanced scams, particularly deepfakes and automated phishing, are increasingly challenging to detect.

4.3 Inferential Analysis

Inferential analysis examines the relationships between variables to test patterns and associations that go beyond mere description. In this study, we explore connections such as internet usage frequency, awareness of scams, prior victimization, and perception of AI involvement in social engineering attacks.

4.3.1 Relationship Between Age and Awareness of Social Media Scams

A cross-tabulation of age groups and awareness shows that younger respondents (18–24) are generally more aware of scams (over 95%) compared to other age brackets. However, older respondents (25–34 and 55+) show slightly lower awareness, confirming that younger students are more active online and likely more exposed to educational content about cyber threats.

Although inferential tests such as the Chi-square test could be used to examine the relationship between age and awareness levels, this study focused primarily on descriptive statistical analysis to present patterns and trends among respondents

4.3.2 Relationship Between Gender and Prior Victimization

Cross-tabulation between gender and prior victimization indicates that 56.4% of respondents overall reported being victims, with slightly more males affected than females.

4.3.3 Frequency of Internet Use vs. Exposure to Scams

Almost all respondents (99.3%) use social media or the internet daily. Due to this high frequency, statistical analysis may be limited in detecting variation, but descriptive comparison shows daily users encounter scams more frequently than those who use the internet less often.

4.3.4 Perception of AI Involvement and Detection Difficulty

Respondents' ability to distinguish AI-generated scams from human scams varies: only 10.7% find it very easy, while 28.6% find it hard. By analyzing this variable against prior victimization or awareness, one can explore whether experience with scams improves the ability to detect AI involvement.

4.3.5 Key Observations from Inferential Analysis

1. Age is positively associated with awareness, with younger students being more exposed to information about scams.
2. Gender does not show a strong difference in prior victimization but requires chi-square testing for confirmation.
3. High internet usage correlates with increased exposure to social engineering attacks, emphasizing the risk for frequent users.
4. Difficulty in detecting AI-generated scams suggests that generative AI enhances the sophistication of attacks, supporting the research objective on AI's role in social engineering.

4.4 Discussion of Key Findings

The analysis of the 140 valid survey responses provides insight into students' awareness, exposure, and perceptions of AI-enhanced social engineering attacks at the University of Benin (UNIBEN). The findings are discussed below in relation to the research objectives and questions.

I. Age and Gender Distribution

The majority of respondents (90%) were aged 18–24 years, reflecting the typical undergraduate population at UNIBEN. Only 9.3% were aged 25–34, and older age groups were almost absent. Gender distribution was fairly balanced, with 52.1% male and 47.9% female respondents. This balance suggests that both male and female students are similarly represented in the study, providing a reliable view of awareness and experiences across genders.

II. Internet and Social Media Usage

Almost all respondents (99.3%) reported using the internet or social media every day. This high frequency of online activity indicates that students are regularly exposed to potential social engineering threats, making them a relevant population for studying AI-driven cyber risks.

III. Awareness of Social Engineering Scams

A significant majority of respondents (95%) were aware of social engineering attacks, which suggests that students have a basic understanding of online threats. This aligns with the research objective of assessing human awareness of AI-enhanced attacks. The 5% of respondents who were unaware indicate that there is still a small group at risk due to lack of knowledge.

IV. Types of Scams Encountered

The survey shows that fake social media accounts (90%), WhatsApp/Telegram scams (87.1%), and job or investment scams (82.9%) were the most common types of attacks students have encountered. Phishing emails and deepfake videos/voice calls were also reported by 68.6% and 65.7% of respondents, respectively. These results demonstrate the wide range of social engineering tactics students face and confirm the literature highlighting social media as a key vector for AI-enhanced attacks (Falade, 2023; Schmitt & Flechais, 2023).

V. Prior Victimization

More than half of respondents (56.4%) reported being victims of online scams, while 38.6% had not experienced any attack, and 5% were unsure. This indicates that exposure to social engineering attacks is common, emphasizing the importance of studying how AI may enhance these threats.

VI. Perception of AI Involvement in Scams

Respondents reported varying abilities to distinguish between human-driven and AI-generated scams. Only 10.7% felt it was very easy, 41.4% somewhat easy, 28.6% hard, 5.7% very hard, and 13.6% were unsure. These results suggest that AI-enhanced social engineering is often difficult for students to detect, highlighting a potential vulnerability that aligns with the research objective of understanding cognitive challenges posed by AI in cybersecurity.

4.4.1 Implications for Research Questions

1. **Emerging tactics and techniques:** The data on phishing, fake accounts, and deepfakes demonstrates that AI is increasingly used to enhance traditional social engineering tactics.
2. **Role of generative AI tools:** Students' difficulties in identifying AI involvement suggest that tools like ChatGPT, FraudGPT, and WormGPT could improve the sophistication of attacks.
3. **Vulnerabilities exploited:** The high rate of prior victimization and difficulty detecting AI-enhanced scams indicate that human cognitive factors remain a key vulnerability.
4. **Effectiveness of current defenses:** Although awareness is high, the continued victimization implies that existing security measures may not be sufficient against AI-powered threats.

4.4.2 Overall Insights

The findings highlight that UNIBEN students are highly exposed to AI-enhanced social engineering, with awareness not always translating into effective protection. This underscores the importance of targeted educational programs, practical cybersecurity training, and the integration of AI-aware defense strategies, echoing the recommendations of Bharati (2024) and Sabatini (2025).

4.5 Implications of the Study

The findings of this research carry significant implications for students, cybersecurity professionals, educational institutions, and policymakers. They highlight both practical and theoretical lessons regarding AI-enhanced social engineering attacks.

1. Implications for Students and Users

The high level of exposure to social engineering attacks, combined with the difficulty in detecting AI-generated scams, suggests that students are at considerable risk online. This underscores the need for stronger personal cybersecurity habits, such as using strong passwords, enabling two-factor authentication, verifying suspicious messages, and staying updated on emerging AI-based threats. Education programs should also focus on helping students recognize AI-driven tactics, not just traditional scams.

2. Implications for Universities and Educational Institutions

Given that the majority of students are active online daily and many have experienced scams, universities have a critical role in reducing vulnerability. Integrating cybersecurity awareness into curricula, offering workshops on AI-enhanced threats, and promoting safe digital behavior are essential steps. UNIBEN, for example, could implement dedicated training sessions on recognizing phishing, deepfakes, and AI-generated scams.

3. Implications for Cybersecurity Professionals

The study confirms that AI-enhanced social engineering is becoming more sophisticated and personalized. Security teams should consider AI-driven defense mechanisms, such as automated threat detection, anomaly monitoring, and human-in-the-loop verification systems. Tools that detect AI-generated content or cross-check multiple communication channels may help reduce successful attacks.

4. Policy and Regulatory Implications

The findings highlight the importance of developing policies and regulations that address AI-driven cyber threats. Policymakers should consider frameworks that enforce safe AI practices, promote transparency, and provide guidance for AI misuse prevention. Ethical guidelines for AI developers could also reduce the risk of these technologies being weaponized against unsuspecting users.

5. Implications for Research and Future Studies

This study provides a baseline for understanding AI-enhanced social engineering in a university context. Future research can expand the sample, examine other Nigerian institutions, or analyze sector-specific vulnerabilities. The difficulty students face in detecting AI-driven scams also points to the need for research into more effective detection tools, public awareness campaigns, and cognitive resilience strategies.

4.5.1 Overall Significance

The study demonstrates that AI-enhanced social engineering is not only a technological problem but also a human and organizational challenge. Addressing it requires a combination of awareness, education, robust technological defenses, and policy measures. By applying these

insights, students, institutions, and security professionals can better prepare for the growing threat posed by AI-driven cyberattacks.

CHAPTER 5

CONCLUSION AND FURTHER DIRECTIONS

5.0 Summary of Key Findings

This study examined the growing influence of artificial intelligence (AI) on social engineering attacks, focusing on awareness, exposure, and perceptions among students of the University of Benin (UNIBEN), Nigeria. Using a descriptive survey design, responses were gathered from 140 participants across different age groups, with the majority (90%) aged between 18 and 24 years. The data provided valuable insights into how young, digitally active populations experience and understand AI-related cybersecurity threats.

The findings revealed that almost all respondents (99.3%) use the internet or social media daily, highlighting the high level of digital engagement among students. This constant connectivity increases exposure to social engineering tactics and emphasizes the importance of digital literacy and cybersecurity awareness.

Regarding awareness of social media scams, an overwhelming 95% of respondents reported being aware of online scams, demonstrating strong general awareness. However, awareness of AI-enhanced scams, such as deepfakes, automated phishing, and AI-generated chatbots appeared less comprehensive, suggesting a gap in understanding the newer, more sophisticated forms of deception.

In terms of types of scams encountered, fake social media accounts (90%), WhatsApp or Telegram scams (87.1%), and job or investment scams (82.9%) were the most reported. This suggests that social engineering attacks in Nigeria are primarily delivered through social networking and messaging platforms rather than email. The data also revealed that 56.4% of respondents had previously fallen victim to some form of online deception, showing that awareness does not necessarily translate into resistance against manipulation.

Perception-based questions further indicated that most respondents find it difficult to distinguish between human and AI-generated scams. Only 10.7% said it was “very easy,” while a combined 34.3% described it as “hard” or “very hard.” This aligns with findings in literature (e.g., Bharati,

2024; Falade, 2023), which emphasize that AI tools significantly enhance the realism and persuasiveness of fraudulent content.

Overall, the findings demonstrate a strong awareness of traditional scams but limited understanding and preparedness for AI-driven deception. The study therefore highlights a growing need for AI-specific cybersecurity education and awareness programs within higher education institutions to address this evolving threat landscape.

5.1 Recommendations

Drawing from the results of this study and aligned with its objective to propose frameworks for mitigating AI-enhanced social engineering attacks, the following recommendations are made:

I. **Integration of Cybersecurity Awareness into Academic Curriculum:** The University of Benin should embed cybersecurity and AI ethics modules into general studies courses for all undergraduate students. These modules should focus on identifying AI-generated scams, understanding manipulation techniques, and practicing secure digital behavior.

II. **Adoption of the Human–AI Cyber Awareness Framework (HACAF):** Based on the findings that 95% of respondents were aware of scams but 56.4% had still experienced one, this research proposes a Human–AI Cyber Awareness Framework (HACAF).

This framework emphasizes:
1. Human Element: Regular awareness campaigns, scenario-based training, and digital hygiene education.
2. AI Element: Deployment of AI-powered monitoring tools that detect phishing, impersonation, and deepfake content.
3. Institutional Integration: A structured feedback mechanism within the university's ICT policy to ensure continuous learning from real incidents.

III. **AI-Powered Scam Detection Tools:**

Students and institutions should integrate AI-based spam filters, identity verification systems, and anomaly detectors to reduce exposure to fake accounts and AI-driven phishing attempts.

IV. **Multi-Level Awareness Campaigns:** Regular workshops should be organized in collaboration with cybersecurity experts to simulate real-world attack scenarios (e.g., fake job offers or AI chatbots) and demonstrate how to identify them.

V. **Institutional Policy and Support Systems:**The university should establish a dedicated Cyber Safety Response Unit (CSRU) under the ICT Directorate to:

1. Handle reports of online fraud and impersonation.
2. Provide immediate response and victim support.
3. Monitor evolving social engineering trends within the student community.

VI. **Public–Private Partnership for Cyber Defense:** Collaboration between the university, private cybersecurity firms, and government agencies (such as NITDA and NCC) should be fostered to build collective resilience against AI-based cyber threats through shared resources, threat intelligence, and capacity-building programs.

By adopting this Human AI Cyber Awareness Framework (HACAF) and institutional support structures, universities can strengthen their human and technological defenses against AI-enhanced social engineering attacks, promoting a safer academic digital environment.

5.2 Suggestions for Future Research

Future studies should expand the scope and depth of inquiry into AI-enhanced social engineering attacks. One key direction would be to include a larger and more diverse sample size that covers students, academic staff, and administrative personnel across different faculties. This would help capture varying levels of exposure and awareness influenced by academic discipline, job role, and digital behavior.

Further research should also aim to apply inferential statistical methods, such as chi-square tests, correlation analysis, or regression models, to establish significant relationships between demographic factors and awareness or victimization levels. This would provide stronger empirical evidence for targeted awareness programs and institutional policies.

In addition, researchers could explore longitudinal studies that track changes in awareness, perception, and attack patterns over time. Since AI technologies evolve rapidly, such studies would help monitor how new attack methods such as voice deepfakes or autonomous scam bots affect user vulnerability.

Future work should also focus on developing and testing AI-based defense frameworks, integrating behavioral analytics, and human-in-the-loop models that can detect manipulation attempts in real time. This aligns with the study's recommendation for proactive, adaptive cybersecurity systems.

Lastly, comparative studies across institutions or regions could provide broader insight into how cultural, infrastructural, and educational differences influence exposure to AI-enabled scams. Such research would enhance both local and global understanding of digital safety in the AI era.

5.3 Conclusion

This study examined the growing impact of AI-enhanced social engineering attacks within the context of the University of Benin, Nigeria. Through a quantitative survey of 140 respondents, primarily aged 18–24, the findings revealed a high level of awareness about social media scams (95%) but also a concerning rate of prior victimization (56.4%). The majority of respondents (90%) identified fake social media accounts and messaging scams as the most common forms of attack, highlighting the prevalence of AI-assisted deception in everyday online environments.

Despite strong awareness levels, the data indicated that many individuals still struggle to identify AI-generated content, as only 52% found it "very easy" or "somewhat easy" to differentiate

between human and AI scammers. This demonstrates a widening awareness–practice gap, where students recognize digital threats but lack the cognitive and technical tools to effectively mitigate them.

This study addressed the following research questions: the emerging tactics and techniques used in AI-enhanced social engineering attacks; the role of generative AI tools such as ChatGPT, FraudGPT, and WormGPT in facilitating sophisticated attacks; the cognitive and technical vulnerabilities exploited by AI-powered attackers; the effectiveness of existing cybersecurity measures at UNIBEN; and practical strategies or frameworks to strengthen institutional resilience. The findings indicate that while awareness is high, technical and behavioral gaps persist, underscoring the need for a structured and integrated approach to cyber defense.

In response, this research proposes the Human–AI Cyber Awareness Framework (HACAF), a hybrid model designed to strengthen digital resilience by aligning human behavioral awareness with AI-powered defense systems. The HACAF model emphasizes three key components:

1. **Human Element**: Building consistent cybersecurity awareness, behavioral training, and digital hygiene habits through experiential learning.

2. **AI Element**: Leveraging artificial intelligence tools for proactive detection of phishing, impersonation, and deepfake manipulation.

3. **Institutional Integration**: Embedding cybersecurity education and reporting structures within university ICT policies for sustainable protection.

By implementing the HACAF model, universities and similar institutions can transition from passive awareness to active cyber resilience, enabling students, educators, and administrators to collaboratively adapt to the evolving threat landscape shaped by generative AI technologies. The study concludes that the future of cybersecurity within academic communities will depend on balancing technological innovation with human vigilance, ensuring that awareness translates into protective action.

REFERENCES

Almutairi, Z., & Elgibreen, H. (2022). *A review of modern audio deepfake detection methods: Challenges and future directions.* IEEE Access, 10, 12234–12257. https://doi.org/10.1109/ACCESS.2022.3147895

Ariza, M., de Almeida, F., Garcia, J., & de Oliveira, L. (2023). *Automated social engineering attacks using chatbots on professional social networks.* Computers & Security, 126, 103056. https://doi.org/10.1016/j.cose.2023.103056

Bharati, R. K. (2024). *AI-enhanced social engineering: Evolving tactics in cyber fraud and manipulation.* Journal of Information Security and Digital Ethics, 13(2), 45–62.

Blauth, T. F., Gstrein, O., & Zwitter, A. (2022). *Artificial intelligence crime: An overview of malicious use and abuse of AI.* AI and Ethics, 2(1), 1–17. https://doi.org/10.1007/s43681-021-00078-9

Falade, P. V. (2023). *Decoding the threat landscape: ChatGPT, FraudGPT, and WormGPT in social engineering attacks.* Cybersecurity Review Journal, 4(2), 67–83.

Gomathi Alias Rohini, S., Rajalakshmi, S., & Nandhini, M. (2025). *Social engineering attacks in cybersecurity.* International Journal of Information Security Research, 11(2), 75–91.

Gupta, M., Badsha, S., Rahman, M. S., & Bertino, E. (2023). *From ChatGPT to ThreatGPT: Impact of generative AI in cybersecurity and privacy.* IEEE Transactions on Technology and Society, 4(3), 200–214. https://doi.org/10.1109/TTS.2023.3298765

Kaur, B. (2025). *Social engineering attacks in the digital age.* International Journal of Cybersecurity Research, 9(1), 22–39.

Kothari, C. R. (2014). *Research methodology: Methods and techniques* (3rd ed.). New Delhi: New Age International.

Kumar, N., & Patel, N. M. (2025). *Social engineering attack in the era of generative AI.* Journal of Emerging Cyber Threats, 5(1), 15–29.

Sabatini, B. M. (2025). *The next generation of cyber threats: AI in social engineering attacks* [Master's dissertation]. University of Essex.

Schmitt, M., & Flechais, I. (2023). *Digital deception: Generative artificial intelligence in social engineering and phishing.* Computers & Security, 128, 103122. https://doi.org/10.1016/j.cose.2023.103122

Yamane, T. (1967). *Statistics: An introductory analysis* (2nd ed.). New York: Harper & Row.

University of Benin. (2025). *About UNIBEN: Facts and figures.* https://www.uniben.edu