

Indian Journal of Modern Research and Reviews

This Journal is a member of the '*Committee on Publication Ethics*'

Online ISSN:2584-184X

Research Article



Adversarial Robustness and Federated Learning Security: A Simulation-Based Analysis of Data Poisoning Defence Systems for Distributed AI in Sub-Saharan Africa

Julia Phillemon ^{1*}, Baljinder Kaur ²

¹ Student, Department of Computer Applications, Guru Kashi University, Talwandi Sabo, Punjab, India

² Assistant Professor, Department of Computer Applications, Guru Kashi University, Talwandi Sabo, Punjab, India

Corresponding Author: *Julia Phillemon

DOI: <https://doi.org/10.5281/zenodo.19941421>

Abstract

This study examines how data poisoning and model poisoning attacks behave in federated learning (FL) environments shaped by the conditions common across Sub-Saharan Africa, with Namibia used as the primary reference context. The motivation for the study came from a straightforward observation: FL frameworks have been proposed as a privacy-preserving solution for distributed AI in low-bandwidth, data-sensitive settings, but the security defences built into them were designed and tested under assumptions that do not hold in most African deployments. This paper draws on a systematic literature review and a set of simulation-based experiments to examine that mismatch directly. Using a Dirichlet-partitioned non-IID environment across 50 simulated clients, the experiments show that standard Byzantine-resilient aggregation methods lose between 26% and 35% of their accuracy performance when IID conditions are replaced with non-IID ones approximating regional African heterogeneity. The study also introduces the FL-STRIDE threat taxonomy, which maps known FL attack vectors onto deployment conditions found in Namibia and similar contexts. A proposed Adaptive Robust Aggregation (ARA) framework, designed with non-IID robustness as a core requirement, achieves an accuracy gap of approximately 11.7% under the same test conditions, though this result is from simulation only and further empirical work is needed. Taken together, the findings suggest that Africa-specific FL security frameworks are both necessary and technically achievable.

Manuscript Information

- ISSN No: 2584-184X
- Received: 04-04-2026
- Accepted: 27-04-2026
- Published: 30-04-2026
- MRR:4(4); 2026: 274-279
- ©2026, All Rights Reserved
- Plagiarism Checked: Yes
- Peer Review Process: Yes

How to Cite this Article

Phillemon J, Kaur B. Adversarial Robustness and Federated Learning Security: A Simulation-Based Analysis of Data Poisoning Defence Systems for Distributed AI in Sub-Saharan Africa. Indian J Mod Res Rev. 2026;4(4):274-279.

Access this Article Online



www.mrrjournal.in

KEYWORDS: Federated Learning, Data Poisoning, Adversarial Machine Learning, Byzantine Robustness, AI Security, Sub-Saharan Africa, Namibia, Non-IID Data.

1. INTRODUCTION

1.1 BACKGROUND AND CONTEXT

AI and machine learning systems are being deployed at growing pace across Sub-Saharan Africa. Healthcare, agriculture, financial services, and public administration have all seen increasing adoption over the past decade. For a long time, African countries were primarily consumers of technology built elsewhere. That picture is changing. Mobile-first infrastructure, expanding start-up ecosystems, and deliberate government investment in digital transformation are creating conditions for locally grounded AI development.

Namibia is a clear example of this direction. The country's Harambee Prosperity Plan II and National Development Plan 5 both identify digital transformation and data-driven governance as strategic priorities. This research study builds directly on earlier work that examined data poisoning risks in traditional ML systems within this context and produced a peer-reviewed analysis of AI and machine learning competencies in cybersecurity roles (Phillemon *et al.*, 2025) ^[10]. The present study extends that work into the domain of federated learning security.

1.2 PROBLEM STATEMENT

Federated learning was introduced by McMahan *et al.* (2017) ^[8] as a way to train shared models across distributed participants without moving raw data to a central server. That property matters a great deal in African settings, where data sovereignty is a genuine concern, bandwidth is unreliable, and devices vary widely in capability. The problem is that FL's distributed architecture also creates its biggest security vulnerability. Because the server cannot inspect what happens inside each client's local training process, the system becomes open to manipulation. Participants who are malicious or compromised can submit poisoned model updates designed to steer the global model in directions that serve the attacker. Bagdasaryan *et al.* (2020) ^[1] showed that these attacks can be highly stealthy and effective even against aggregation methods that were designed to be robust.

What has not been addressed in the existing literature is how these attacks and the defences against them behave specifically in the non-IID, low-bandwidth, under-regulated conditions that characterise African FL deployments. That gap is what this study addresses. The central research question guiding the work is:

How do data poisoning and model poisoning attacks affect federated learning performance under the non-IID, resource-constrained conditions characteristic of Sub-Saharan African deployments, and what defence approaches may be better suited to these settings?

1.3 Scope and Structure

The scope of this study is deliberately focused. Rather than attempting to cover all of Africa or all possible FL deployment scenarios, the study uses Namibia as a grounded reference case and employs simulation to test attack and defence behaviour under conditions informed by that context. Section 2 reviews prior work across three relevant areas. Section 3 sets out the

theoretical framework and the proposed ARA architecture. Section 4 describes the research methodology in detail. Section 5 presents the experimental results and discusses what they show. Section 6 concludes and identifies directions for future research.

2. LITERATURE REVIEW

This section reviews prior work across three connected areas: data poisoning in traditional machine learning, the security vulnerabilities of federated learning architectures, and the specific conditions of AI deployment in Sub-Saharan Africa. The goal is to show where existing knowledge sits and what it leaves unresolved.

2.1 Data Poisoning in Traditional Machine Learning

The study of adversarial attacks on machine learning systems has a reasonably well-developed foundation. Biggio *et al.* (2012) ^[2] were among the first to show that a small number of carefully crafted training examples could degrade the performance of support vector machine classifiers, and their work established the important conceptual distinction between poisoning attacks, which operate on training data, and evasion attacks, which manipulate inputs at inference time. Gu *et al.* (2017) ^[6] extended this line of work into deep learning with the BadNets framework, demonstrating that backdoors could be embedded in a model so that it behaved correctly on ordinary inputs but produced attacker-chosen outputs whenever a specific trigger pattern appeared. Shafahi *et al.* (2018) ^[11] pushed the attack surface further with clean-label poisoning, a technique in which the adversary manipulates the feature space of training examples without altering their labels, making detection substantially harder because the data looks correct to a human reviewer.

Phillemon *et al.* (2025) ^[10] synthesised and extended these frameworks within the Namibian context, developing a risk-scoring model for organisational ML deployments that quantified poisoning exposure as a function of data provenance controls, training pipeline visibility, model interpretability, and the criticality of the application. That earlier work forms the direct starting point for the present study, which takes the same risk orientation and applies it to federated architectures.

2.2 Federated Learning Architecture and Its Security Vulnerabilities

McMahan *et al.* (2017) ^[8] introduced FedAvg, the foundational FL protocol, in which each participating client trains a local model update and submits it to a central server for aggregation into a global model. Raw data never leaves the client, which is a meaningful practical advantage where data sharing is legally constrained or technically difficult. The core security problem is that FL's architecture removes the server's ability to verify what happened during local training. A client that is compromised or acting in bad faith can submit gradient updates that look similar to legitimate ones but that push the global model in a direction chosen by the attacker.

Several Byzantine-resilient aggregation methods have been proposed to address this. Blanchard *et al.* (2017) ^[3] introduced

Krum, which selects the submitted update that is geometrically closest to its neighbours. Yin *et al.* (2018) ^[13] proposed coordinate-wise Trimmed Mean, which clips the most extreme values in each coordinate before aggregating. Cao *et al.* (2020) ^[4] developed FLTrust, which uses a small server-held trusted dataset as a reference for scoring client updates. All three methods have demonstrated effectiveness under IID data conditions. However, as the simulation results in Section 5 show, their performance degrades substantially when data is distributed in the non-IID patterns that are normal in African settings. This limitation has not been systematically addressed in the existing literature.

2.3 AI Deployment Conditions in Sub-Saharan Africa

The literature connecting AI security to African digital development is still limited. Gwagwa *et al.* (2022) ^[7] provided one of the more comprehensive analyses of AI in Africa, identifying structural barriers to trustworthy implementation. Cloete (2021) ^[5] examined Namibia in particular, finding significant gaps in data governance frameworks and noting the absence of clear institutional guidance on security requirements for AI systems. These findings matter for FL security because the threat environment is shaped by institutional as much as technical factors.

Four features of the African setting are particularly relevant to the security problem. First, data distributions across federated participants are highly non-IID because of demographic, linguistic, socioeconomic, and geographic heterogeneity. This directly violates the distributional assumptions underpinning most existing Byzantine-resilient methods. Second, network

connectivity is intermittent and variable, meaning that clients drop in and out of training rounds unpredictably. Third, regulatory and institutional oversight is limited, which reduces the deterrent effect on malicious participants and leaves deploying organisations without adequate guidance. Fourth, the applications being targeted are high-stakes: agricultural advisory systems, credit scoring for mobile finance, and public health surveillance are all contexts where a compromised model can cause real harm (Ajuzieogu, 2021; Okolo *et al.*, 2024) ^[14, 15].

3. Theoretical and Conceptual Framework

This section presents the two main conceptual contributions of the study: the FL-STRIDE threat taxonomy and the proposed Adaptive Robust Aggregation architecture.

3.1 The FL-STRIDE Threat Taxonomy

Threat modelling is a structured way of identifying what an adversary might do, why, and where in a system they might do it. The STRIDE framework, originally developed for distributed software systems, organises threats into six categories: Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, and Elevation of Privilege. This study adapts that structure to the federated learning threat surface, using Namibian and African deployment scenarios as grounding examples. The result is the FL-STRIDE taxonomy presented in Table 1. The value of this taxonomy is not that it is exhaustive but that it forces clarity about attacker goals rather than just attack techniques, which tends to produce more useful security analysis.

Table 1: FL-STRIDE Threat Taxonomy with African Deployment Context Examples

Stride Category	FL Attack Vector	Example in Namibian / African Context
Spoofing	Client identity fraud; Sybil attacks	A mobile money operator registers multiple fraudulent client identities to amplify the poisoning impact on a shared credit model.
Tampering	Data poisoning; Gradient manipulation	A compromised agricultural IoT sensor injects corrupted crop-yield readings into local FL training rounds, degrading harvest advisory predictions.
Repudiation	Untraceable gradient attacks	A malicious client exploits the absence of audit logging to submit poisoned updates with no traceable attribution.
Information Disclosure	Gradient inversion; Inference attacks	A server-side adversary reconstructs sensitive patient records from gradient updates submitted by clinic nodes in a federated health surveillance system.
Denial of Service	Slow learner attack; Byzantine flooding	An adversary deliberately submits computationally expensive or conflicting model updates to stall training convergence across the federation.
Elevation of Privilege	Model replacement; Backdoor injection	A fintech adversary hijacks a shared credit-scoring model to approve fraudulent loan applications by embedding a trigger-activated backdoor.

Source: Authors' adaptation of the STRIDE threat modelling framework (Shostack, 2014) ^[12] to federated learning attack surfaces.

3.2 The Proposed Adaptive Robust Aggregation (ARA) Framework

The existing Byzantine-resilient aggregation methods reviewed in Section 2.2 each address part of the problem, but none were designed with non-IID data or resource-constrained environments as primary requirements. The ARA framework proposed in this study is intended to address those gaps through a layered defence approach. The core principle is that relying on a single mechanism is insufficient because sophisticated adversaries can adapt to any fixed defence. Resilience, instead, needs to come from layering complementary controls at different points in the FL pipeline.

The proposed architecture has three layers. The first is a client-side integrity layer, which applies local differential privacy mechanisms and data provenance controls before any update is submitted. The intent is to limit how much information a malicious gradient update can carry, regardless of what the client does locally. The second is an aggregation-side robustness layer, which combines reputation-weighted update scoring, spectral outlier detection, and adaptive clipping calibrated for non-IID data. This layer is the primary extension beyond existing methods. The third is a system-level monitoring layer, which tracks model output distributions after

each aggregation round and triggers rollback protocols if anomalies consistent with adversarial compromise are detected. The ARA design described here is a proposed architecture, not a finished system. The simulation results in Section 5 offer early evidence of its potential, but the framework has not been tested in any real deployment. It is put forward as a structured basis for further empirical work, and the claims made for it should be read in that light.

4. RESEARCH METHODOLOGY

This section describes how the study was conducted, covering research design, literature review methodology, and simulation design.

4.1 Research Design

The study uses a mixed-methods design that combines a systematic literature review with simulation-based experimentation. The two methods serve different purposes. The literature review locates the study within existing knowledge, identifies the specific gap being addressed, and provides the conceptual grounding for the threat taxonomy and the ARA framework. The simulation experiments then test specific empirical claims about how existing defences perform under non-IID conditions and how the proposed approach compares. This combination was chosen because a purely theoretical approach would not produce the kind of concrete performance evidence that FL practitioners and researchers need, while a purely technical approach without the literature grounding would struggle to situate the findings in a meaningful way.

4.2 LITERATURE REVIEW METHODOLOGY

The literature review followed a PRISMA-guided search protocol covering FL security, Byzantine-resilient aggregation, adversarial machine learning, and AI deployment in resource-constrained environments. The search covered publications from 2012 to 2025, with particular weight given to work published after 2017 when federated learning became a mainstream research area. Approximately 200 papers were initially screened. The inclusion criteria required papers to engage with FL security or Byzantine robustness, or to address AI deployment conditions in developing-country or low-infrastructure contexts. Papers addressing only centralised ML security without applicability to distributed settings were excluded. Between 60 and 80 papers were carried out in full synthesis.

4.3 SIMULATION DESIGN

The simulation experiments used the MNIST dataset as the benchmark classification task. MNIST was chosen for

consistency with prior FL security studies (Bagdasaryan *et al.*, 2020; Cao *et al.*, 2020) ^[1, 4], enabling direct comparison of baseline results. The dataset was partitioned across 50 simulated clients using a Dirichlet distribution with concentration parameter alpha equal to 0.5, a standard approach for approximating the label heterogeneity that arises from real-world non-IID data. The Dirichlet value of 0.5 produces moderately to heavily skewed distributions, which is consistent with the demographic and geographic heterogeneity described in the African AI literature.

Training ran for 100 communication rounds. Byzantine client ratios were tested at four levels: 5%, 10%, 20%, and 30% of the total client pool. Six attack types were evaluated across these conditions: label flipping, clean-label poisoning, backdoor injection, model replacement, gradient inversion, and slow learner attacks. The proposed ARA approach was implemented and evaluated under the same conditions as the baselines. All simulations were run in Python using the Flower (flwr) FL framework. Each reported figure is an average across five independent runs; observed variance across runs was low, with standard deviations below 1.2 percentage points for all methods, suggesting the results are stable rather than artefacts of a particular random seed.

The primary performance metric was accuracy retention, defined as the global model accuracy under attack as a percentage of clean-training baseline accuracy. Secondary metrics were the attacking success rate for backdoor-type attacks and computational overhead relative to FedAvg.

4.4 Ethical Considerations

This study used only publicly available benchmark datasets and did not involve human participants. The threat taxonomy and defence framework are oriented entirely toward defensive and protective purposes. All vulnerability-related discussion is framed in terms of systemic security analysis consistent with responsible disclosure norms.

5. RESULTS AND DISCUSSION

This section presents the simulation results and discusses what they show, both in terms of the specific findings and their broader implications for FL security in African settings.

5.1 Performance Degradation under Non-IID Conditions

The most consistent finding across the experiments is that all four baseline aggregation methods experience a substantial drop in accuracy retention when the data distribution shifts from IID to non-IID. Table 2 summarises these results for the 20% Byzantine client scenario under a label-flipping attack, which produced the most interpretable comparisons across methods.

Table 2: Accuracy Retention under IID vs. Non-IID Conditions (20% Byzantine Clients, Label-Flipping Attack)

Aggregation Method	Accuracy (IID)	Accuracy (Non-IID, Simulated)	Performance Gap
FedAvg	81.2%	46.5%	34.7%
Krum	89.0%	61.0%	28.0%
Trimmed Mean	87.5%	58.3%	29.2%
FLTrust	91.2%	64.7%	26.5%
Proposed ARA (Simulated)	90.1%	78.4%	11.7%

Note: ARA results are from simulation only and have not been validated in a live deployment. IID baseline figures for Krum and Trimmed Mean are referenced from Blanchard *et al.* (2017) ^[3] and Yin *et al.* (2018) ^[13], respectively.

FedAvg, which makes no claim to Byzantine resilience, dropped 34.7 percentage points, confirming that the threat is severe under these conditions. More telling is the behaviour of methods explicitly designed for robustness. Krum dropped from 89.0% under IID to 61.0% under non-IID, a gap of 28 percentage points. FLTrust, the strongest performer among the baselines, still lost 26.5 percentage points. The common factor across all of these is that their robustness guarantees depend on distributional assumptions that break down when clients hold genuinely different data.

The proposed ARA approach achieved 78.4% accuracy retention under the same non-IID conditions, a gap of 11.7 percentage points from its IID performance. While this is still a meaningful drop, it represents a substantially smaller degradation than any of the baselines.

That difference is attributable primarily to the adaptive clipping and reputation-weighted scoring in the aggregation layer, which are calibrated to tolerate skewed update distributions rather than treat deviation from a global distribution as a signal of adversarial behaviour.

5.2 Backdoor Attack Resistance

Model replacement attacks presented a particular challenge for all baseline methods. Under Trimmed Mean aggregation, a backdoor was successfully embedded with a 97.3% trigger-activation rate when the adversary constrained the magnitude of submitted updates to fall within normal-looking ranges. This is consistent with the findings of Bagdasaryan *et al.* (2020) ^[1] in IID settings, and the results here suggest the problem is at least as severe under non-IID conditions. The ARA approach's system-level monitoring layer, which tracks output distribution anomalies across rounds, reduced backdoor persistence in the simulation, though this result requires further testing under adaptive adversary conditions before strong claims can be made.

5.3 Qualitative Comparison of Defence Mechanisms

Table 3 provides a structured qualitative comparison of the approaches evaluated in this study across the dimensions most relevant to African deployment contexts. The table is intended to support practical decision-making rather than to make absolute performance claims.

Table 3: Qualitative Comparison of FL Defence Mechanisms for African Deployment Contexts

Defence Method	Non-IID Robust	Low-Bandwidth Suitable	Backdoor Resistant	Computational Overhead
FedAvg	No	Yes	No	Low
Krum	Partial	Yes	Partial	Medium
Trimmed Mean	Partial	Yes	Partial	Medium
FLTrust	Partial	Yes	Yes	Medium
Proposed ARA	Yes (by design)	Yes	Yes (by design)	Medium-High

Note: ARA properties marked 'by design' reflect architectural intent confirmed in simulation. Full empirical validation is required before these properties can be treated as guarantees.

5.4 Threat Surface Assessment: Namibian FL Deployments

Drawing on the FL-STRIDE taxonomy and the simulation results together, three Namibian deployment contexts emerge as particularly high-risk from a federated learning security standpoint. Agricultural advisory systems, which would aggregate crop health data from farmer cooperatives across ecologically diverse regions, face intrinsically non-IID data distributions and intermittent connectivity, conditions under which the standard defences have been shown to fail. Credit scoring in mobile financial services presents a different risk profile: adversarial participants in this context have direct financial incentives for targeted manipulation, and the consequences of a compromised credit model are immediate and concrete. Public health surveillance using federated clinic data faces both a severe privacy constraint, making gradient inversion attacks particularly dangerous, and an infrastructure constraint that makes richer trust-bootstrapping mechanisms difficult to implement.

5.5 LIMITATIONS

There are several things this study cannot claim. The most significant limitation is that all performance results are from simulation. The Dirichlet non-IID partitioning is a reasonable approximation of heterogeneous data distributions, but it does not capture real-world complexity, including temporal shifts, device dropouts, or the strategic behaviour of human adversaries who adapt their attacks as defences change. The benchmark dataset used, MNIST, is a well-understood but relatively simple image classification problem. Results on tabular datasets more representative of African FL use cases, such as agricultural yield data or mobile transaction records, could differ in important ways. The ARA framework has not been subjected to adversarial evaluation by independent researchers, which limits confidence in its robustness claims. Future work should address these limitations through empirical pilots in partnership with Namibian organisations, and through evaluation on domain-relevant datasets.

6. CONCLUSION

This study set out to investigate how data poisoning and model poisoning attacks affect federated learning systems under conditions shaped by Sub-Saharan African deployments, and whether defence approaches designed for those conditions could perform better than existing methods. The simulation results support two conclusions. First, the standard Byzantine-resilient aggregation methods available today lose a substantial portion of their protective effect when data is distributed in the non-IID patterns that are normal across African FL participants, with accuracy gaps of 26 to 35 percentage points relative to IID performance. Second, the proposed ARA framework, designed with non-IID conditions as a core requirement, achieved a meaningfully smaller gap of 11.7 percentage points in the same simulation environment.

The FL-STRIDE threat taxonomy developed as part of this study offers a structured way to think about FL security risks in African deployment contexts, grounding abstract attack categories in concrete scenarios relevant to mobile finance, agricultural advisory, and public health. These contributions, taken together, are intended as a foundation for further work rather than as definitive results. The simulation environment has limitations that only empirical testing in real deployments can address. What the study does establish is that the gap identified in the literature is real and technically meaningful, and that Africa-specific FL security frameworks are a worthwhile and achievable direction for future research.

REFERENCES

1. Bagdasaryan E, Veit A, Hua Y, Estrin D, Shmatikov V. How to backdoor federated learning. In: Proc 23rd Int Conf Artificial Intelligence and Statistics (AISTATS); 2020. PMLR Vol. 108, pp. 2938–2948.
2. Biggio B, Nelson B, Laskov P. Poisoning attacks against support vector machines. In: Proc 29th Int Conf Machine Learning (ICML); 2012. Edinburgh, pp. 1807–1814.
3. Blanchard P, El Mhamdi EM, Guerraoui R, Stainer J. Machine learning with adversaries: Byzantine tolerant gradient descent. In: Adv Neural Inf Process Syst (NeurIPS); 2017. Vol. 30.
4. Cao X, Fang M, Liu J, Gong NZ. FLTrust: Byzantine-robust federated learning via trust bootstrapping. arXiv [Preprint]. 2020;2012.13995.
5. Cloete L. Digital governance and data sovereignty in Namibia: challenges and emerging frameworks. Afr J Inf Commun. 2021;27:1–22.
6. Gu T, Dolan-Gavitt B, Garg S. BadNets: identifying vulnerabilities in the machine learning model supply chain. arXiv [Preprint]. 2017;1708.06733.
7. Gwagwa A, Kazim E, Hilliard A, Arvanitis G. The role of artificial intelligence in Sub-Saharan Africa: challenges and opportunities. AI Ethics. 2022;2(3):455–476.
8. McMahan B, Moore E, Ramage D, Hampson S, Arcas BA. Communication-efficient learning of deep networks from decentralised data. In: Proc 20th Int Conf Artificial Intelligence and Statistics (AISTATS); 2017. PMLR Vol. 54, pp. 1273–1282.
9. Ministry of Information and Communication Technology. National broadband policy and digital transformation roadmap 2022–2027. Windhoek: Government of the Republic of Namibia; 2022.
10. Phillemon JAN, Gupta S, Rani S. A literature review of AI and machine learning competencies in modern cybersecurity roles. Indian J Mod Res Rev. 2025;3(11):53–57. doi:10.5281/zenodo.17825894.
11. Shafahi A, Huang WR, Najibi M, Suci O, Studer C, Dumitras T, et al. Poison frogs! Targeted clean-label poisoning attacks on neural networks. In: Adv Neural Inf Process Syst (NeurIPS); 2018. Vol. 31.
12. Shostack A. Threat modeling: designing for security. Hoboken (NJ): John Wiley & Sons; 2014.
13. Yin D, Chen Y, Kannan R, Bartlett P. Byzantine-robust distributed learning: towards optimal statistical rates. In: Proc 35th Int Conf Machine Learning (ICML); 2018. PMLR.
14. Ajuzieogu UC. Federated learning and privacy-preserving generative AI for African contexts: a research approach. ResearchGate [Preprint]. 2021.
15. Okolo JN, Ibiyeye AO, Adim E, Adeniji SA. An extensive review of artificial intelligence utilisation in data science for strengthened cybersecurity analytics, predictive threat assessment and advanced risk management strategies. Appl Sci Comput Energy. 2024;1(1):206–230.

Creative Commons License

This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution–Non-commercial–No Derivatives 4.0 International (CC BY-NC-ND 4.0) License. This license permits users to copy and redistribute the material in any medium or format for non-commercial purposes only, provided that appropriate credit is given to the original author(s) and the source. No modifications, adaptations, or derivative works are permitted.

About the author



Julia Phillemon is a Namibian researcher completing her MSc in Information Technology at Guru Kashi University, India. Her work focuses on AI security, federated learning and responsible AI deployment in African contexts. She has two peer-reviewed publications and practical experience in IT support and data analytics across Namibia and India.



Baljinder Kaur is an Assistant Professor in the Department of Computer Applications at Guru Kashi University, Talwandi Sabo, Punjab, India. She specialises in computer science, software development, and information technology, contributing to teaching, research, and academic development while guiding students in emerging technological advancements and practical applications.