COMPUTING AND ARTIFICIAL INTELLIGENCE FOR FUTURE NETWORKS

INFRASTRUCTURE BARRIERS TO AI IMPLEMENTATION AND THE STRATEGIC ROLES OF TELECOMMUNICATIONS AND AFRICAN GOVERNMENTS AS CO-ARCHITECTS

Jackson Phiri (PhD)
Associate Professor – Department of Computer Science
The University of Zambia

In this Presentation, we argue that the AI-enabled future network is not a distant vision for Africa and other developing countries, it is a partnership problem solvable with existing assets

PRESENTATION OUTLINE

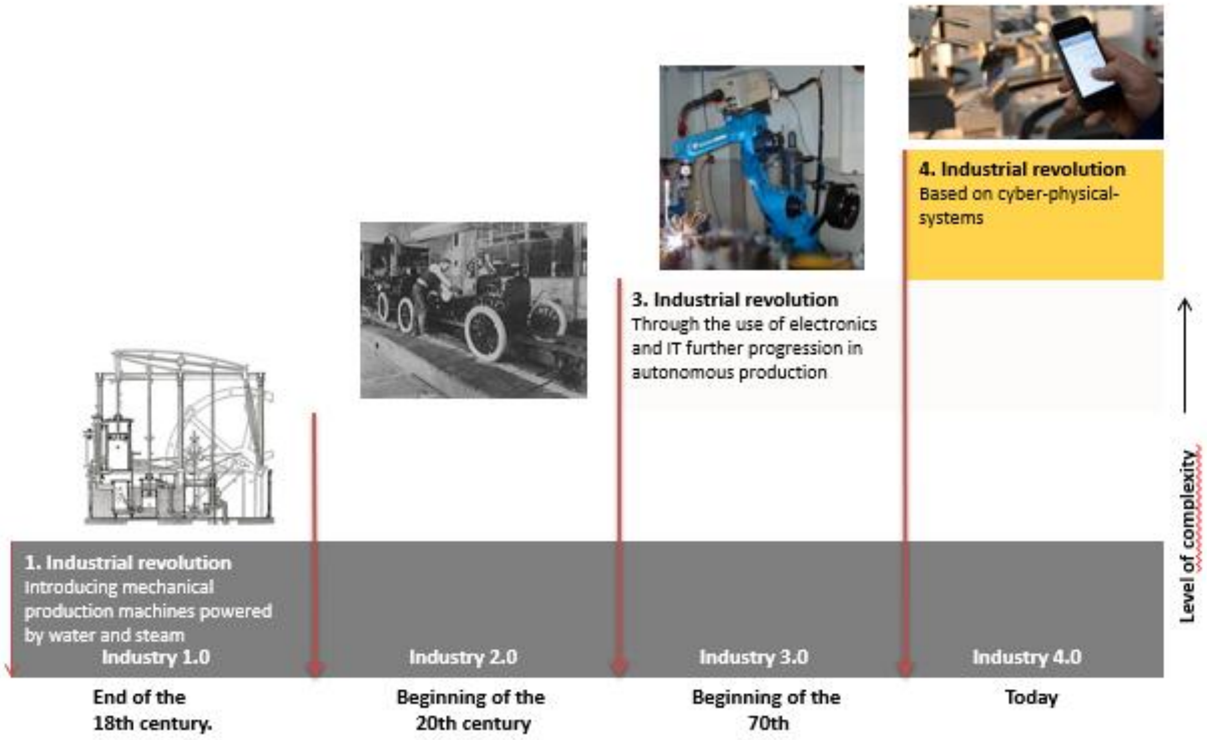
- Introduction
- Problem of the the Infrastructure Barriers
- Proposal for AI implementation to overcome the Infrastructure Barriers in Africa and developing countries
  - Governments and Telecommunication Companies Strategies
  - Distributed Artificial Intelligence Control Plane
  - Federated Learning For Privacy
- Recommendations
- Use Cases Examples
- References

INTRODUCTION: THE NEW DIGITAL ERA

Source: https://www.unido.org/unido-industry-40

INTRODUCTION: THE NEW DIGITAL ERA (Industry 4.0)

- According to the United Nations Industrial Development Organisation, we are now entering in the New Digital Era in this century
  - https://www.unido.org/unido-industry-40
  - This Era is generally Referred to as **Fourth Industrial Revolution** (Industry 4.0)
- The concept of the Fourth Industrial Revolution was coined in 2016 by Klaus Schwab, the founder of the World Economic Forum



Source: DFKI/Bauer IAO

Source: https://www.weforum.org/publications/top-10-emerging-technologies-of-2023/

APR 2026

ESSENTIAL DIGITAL HEADLINES

OVERVIEW OF THE ADOPTION AND USE OF CONNECTED DEVICES AND SERVICES

Source: https://datareportal.com/reports/digital-2026-mid-year-global-update-report

GLOBAL OVERVIEW

TOTAL POPULATION	UNIQUE MOBILE PHONE SUBSCRIBERS	INDIVIDUALS USING THE INTERNET	SOCIAL MEDIA USER IDENTITIES	MONTHLY ACTIVE GENERATIVE AI USERS
8.28 BILLION	**5.83 BILLION**	**6.12 BILLION**	**5.79 BILLION**	**2.42 BILLION**
URBANISATION	vs. POPULATION	vs. POPULATION	vs. POPULATION	vs. POPULATION
80.5%	**70.4%**	**73.8%**	**69.9%**	**29.2%**

SOURCES: MANOCHI ANALYSIS; UNITED NATIONS; GOVERNMENT RESOURCES; GSMA INTELLIGENCE; ITU; EUROSTAT; CNNIC; SIMILARWEB; PLATFORM RESOURCES; COMPANY EARNINGS REPORTS; OCDH; BETA RESEARCH CENTER. **ADVISORY:** SOCIAL MEDIA USER IDENTITIES AND GENERATIVE AI USERS MAY NOT REPRESENT UNIQUE INDIVIDUALS. **NOTES:** WHERE AVAILABLE, THE FIGURE FOR URBANISATION REPRESENTS THE U.N.'S REPORTED VALUE FOR THE PERCENTAGE OF THE TOTAL POPULATION LIVING IN TOWNS AND CITIES. FIGURES FOR GENERATIVE AI USE ONLY INCLUDE THE USE OF STANDALONE GENERATIVE AI PLATFORMS, AND DO NOT FACTOR THE USE OF AI FUNCTIONALITY WITHIN OTHER SOFTWARE OR APPLICATIONS. **COMPARABILITY:** BASE REVISIONS; SOURCE CHANGES. SEE NOTES ON DATA.

we are social MANOCHI

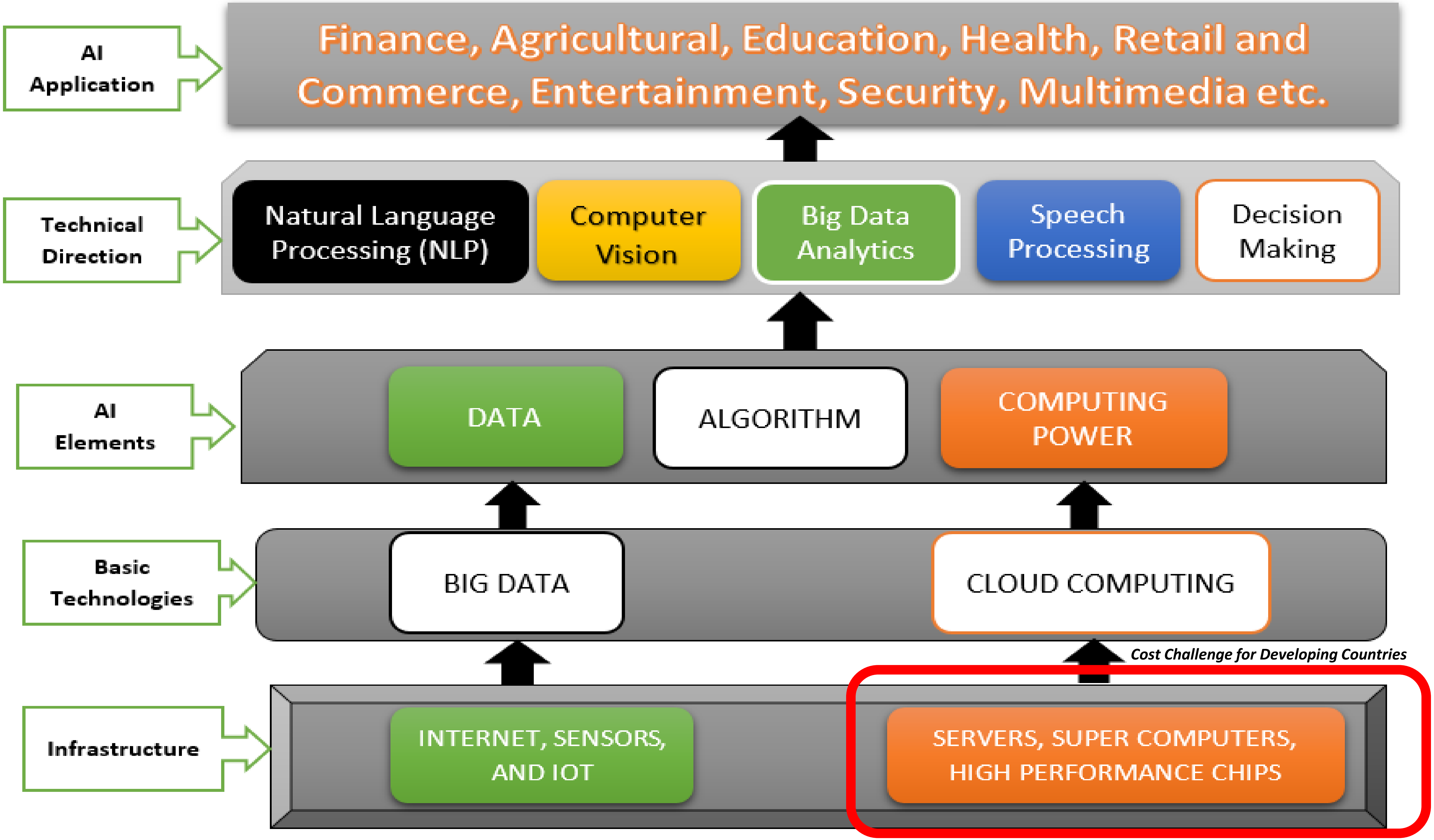
PROBLEM

INFRASTRUCTURE BARRIERS TO AI IMPLEMENTATION IN AFRICA

- **The Core Argument**
  - Future networks (6G, O-RAN) are AI native and they require distributed intelligence.
  - Africa's infrastructure gaps (**POWER**, **INTERNET CONECTIVITY**, **COMPUTE**) make cloud dependent AI impossible.
  - Telecommunication companies and governments together already own the pieces, they just need to connect them.
- ***CONCLUSION: This is a coordination problem.***

Infrastructure Barriers – The Triad (Africa-Specific)

Barrier	Impact on AI for Future Networks
Unreliable grid	AI servers at towers cannot run 24/7; diesel is expensive and unsustainable
Low backhaul capacity	Sending telemetry to central cloud fails; distributed architectures required
No local compute	No edge inference → RAN Intelligent Controller (RIC) impossible
High latency to foreign clouds	>200ms kills real-time network optimisation
Import tariffs on AI hardware	GPUs, edge accelerators cost 2–3× global price, stalling deployment
Fragmented regulation	Cross-border AI models (e.g., roaming optimisation) illegal or impractical


AI Application
Finance, Agricultural, Education, Health, Retail and Commerce, Entertainment, Security, Multimedia etc.
Technical Direction
Natural Language Processing (NLP)
Computer Vision
Big Data Analytics
Speech Processing
Decision Making
AI Elements
DATA
ALGORITHM
COMPUTING POWER
Basic Technologies
BIG DATA
CLOUD COMPUTING
Cost Challenge for Developing Countries
Infrastructure
INTERNET, SENSORS, AND IOT
SERVERS, SUPER COMPUTERS, HIGH PERFORMANCE CHIPS

KEY GPUs AND AI ACCELERATORS USED IN AI DATA CENTER INFRASTRUCTURE

MANUFACTURER	MODEL / SERIES	KEY SPECS & USE CASE
NVIDIA	H100 (Hopper)	640GB GPU memory, 700W TDP; industry standard for AI training & inference
	H200 (Hopper)	76% more memory & 43% higher bandwidth than H100; ideal for LLMs
	B200 (Blackwell)	1000W TDP; up to 3x faster training, 15x faster inference than H100
	GB200 (Grace Blackwell)	Superchip with Grace CPU + 2 Blackwell GPUs; for massive AI-scale workloads
	GB300 NVL72 (Blackwell Ultra)	Rack-scale: 72 GPUs + 36 CPUs; extreme LLM training & inference
AMD	Instinct MI300X	1.5TB HBM3 memory, 5.3 TB/s bandwidth; generative AI & HPC
	Instinct MI350X	Upcoming series for next-gen AI data centers
Intel	Gaudi 3	128GB HBM per GPU, 8-GPU system bandwidth 3.7 TB/s; cost-efficient training & inference
	Crescent Island (Xe3P)	New data center GPU optimized for AI inference
Emerging / Alternative	Cerebras WSE-3	Wafer-scale engine, 125 PFLOPS; trains massive models without complex clusters
	Google TPU, Amazon Trainium/Inferentia, Groq LPU, Graphcore IPU, Positron Atlas	Custom or specialized chips tailored for specific cloud workloads, low latency, or unique architectures

Rank	Model	Performance Notes	Hard Price (USD)
1	GB300 NVL72 (NVIDIA)	Rack-scale; 72x Blackwell Ultra GPUs	~4.3M–4.3M–6.5 M per rack
2	GB200 (NVIDIA)	Superchip (2x Blackwell GPUs + Grace CPU)	~60K–60K–70K per superchip
3	B200 (NVIDIA)	3x training, 15x inference vs H100	35K–35K–40K
4	Cerebras WSE-3	Wafer-scale; 125 PFLOPS	>$2M per system (est.)
5	H200 (NVIDIA)	Higher memory bandwidth vs H100	27K–27K–45K
6	H100 (NVIDIA)	Industry baseline	25K–25K–40K
7	MI300X (AMD)	192GB HBM3 memory	10K–10K–20K
8	Gaudi 3 (Intel)	Cost-efficient AI training	~$15K per accelerator
9	MI350X (AMD)	Upcoming, 288GB HBM3e	N/A (Not yet available)

PROBLEM

INFRASTRUCTURE BARRIERS TO AI IMPLEMENTATION IN AFRICA

- **Scientific Publication:** Gwagwa, Kazim and Kachali (2023)
  - There is a concern that the digital divide in Africa presents significant challenges for the adoption of AI.
  - The digital divide is rooted in issues such as inadequate infrastructure, unaffordability, and limited digital skills.
- **Scientific Publication:** Das, Farid, and Saha (2025)
  - Over 80% of the world's population resides in emerging countries where telecommunications infrastructure remains a critical constraint on AI adoption.
- ***CONCLUSION***: *These are design constraints, not minor hurdles. Build around them or fail.*

Infrastructure Barriers – The Triad (Africa-Specific)

Barrier	Impact on AI for Future Networks
Unreliable grid	AI servers at towers cannot run 24/7; diesel is expensive and unsustainable
Low backhaul capacity	Sending telemetry to central cloud fails; distributed architectures required
No local compute	No edge inference → RAN Intelligent Controller (RIC) impossible
High latency to foreign clouds	>200ms kills real-time network optimisation
Import tariffs on AI hardware	GPUs, edge accelerators cost 2–3× global price, stalling deployment
Fragmented regulation	Cross-border AI models (e.g., roaming optimisation) illegal or impractical

THE DIGITAL DIVIDE AND INTERNET ACCESS

APR 2026

WORLD'S LARGEST CONNECTED POPULATIONS

COUNTRIES WITH THE GREATEST NUMBER OF PEOPLE USING THE INTERNET

GLOBAL OVERVIEW

#	COUNTRY	USERS	vs. POP	ΔYOY
01	CHINA	1,297,764,000	91.8%	-0.4%
02	INDIA	1,031,420,000	70.0%	+0.9%
03	UNITED STATES OF AMERICA	330,109,000	94.7%	+0.5%
04	INDONESIA	230,285,000	80.1%	+3.1%
05	BRAZIL	184,469,000	86.5%	+0.1%
06	PAKISTAN	147,865,000	57.3%	+1.6%
07	RUSSIAN FEDERATION	135,392,000	94.4%	-0.5%
08	NIGERIA	112,566,000	46.7%	+4.3%
09	MEXICO	110,335,000	83.1%	+0.4%
10	JAPAN	104,874,000	85.5%	-0.5%
11	EGYPT	98,653,000	82.4%	+2.2%
12	PHILIPPINES	95,309,000	81.1%	+3.7%
13	BANGLADESH	94,709,000	53.4%	+1.2%
14	VIETNAM	85,864,000	84.2%	+0.6%
15	GERMANY	79,960,000	95.5%	-0.7%
16	TURKEY	79,417,000	90.4%	+0.3%
17	IRAN	79,359,000	85.3%	+0.8%
18	UNITED KINGDOM	68,276,000	97.8%	+0.6%
19	THAILAND	67,675,000	94.6%	+0.2%
20	FRANCE	63,848,000	95.7%	+0.1%

46 **SOURCES:** MANOCHI ANALYSIS; ITU; GSMA INTELLIGENCE; EUROSTAT; GOOGLE'S ADVERTISING RESOURCES; CNNIC; LOCAL GOVERNMENT AUTHORITIES; UNITED NATIONS. **NOTE:** VALUES HAVE BEEN CAPPED AT 99% OF THE POPULATION. **COMPARABILITY:** SOURCE AND BASE CHANGES. ALL FIGURES USE THE LATEST AVAILABLE DATA, BUT SOME SOURCES DO NOT PUBLISH REGULAR UPDATES, SO FIGURES MAY UNDER-REPRESENT ACTUAL USE. SEE NOTES ON DATA.

we are social MANOCHI

APR 2026

UNCONNECTED POPULATIONS

COUNTRIES AND TERRITORIES WITH THE LARGEST UNCONNECTED POPULATIONS AND THE LOWEST LEVELS OF INTERNET ADOPTION

GLOBAL OVERVIEW

ABSOLUTE: LARGEST UNCONNECTED POPULATIONS

#	LOCATION	UNCONNECTED INDIVIDUALS	% OF POP. OFFLINE
01	INDIA	442,037,000	30.0%
02	NIGERIA	128,630,000	53.3%
03	CHINA	115,922,000	8.2%
04	PAKISTAN	110,401,000	42.7%
05	ETHIOPIA	107,133,000	77.6%
06	DEM. REP. OF THE CONGO	92,807,000	80.3%
07	BANGLADESH	82,577,000	46.6%
08	INDONESIA	57,068,000	19.9%
09	TANZANIA	49,601,000	68.8%
10	UGANDA	40,796,000	77.8%

RELATIVE: LOWEST LEVELS OF INTERNET ADOPTION

#	LOCATION	% OF POP. OFFLINE	UNCONNECTED INDIVIDUALS
01	NORTH KOREA[1]	>99%	[BLOCKED]
02	BURUNDI	89.9%	13,159,000
03	CHAD	87.4%	18,715,000
04	SOUTH SUDAN	86.5%	10,706,000
05	CENTRAL AFRICAN REPUBLIC	86.2%	4,873,000
06	NIGER	83.0%	23,733,000
07	YEMEN	82.3%	35,115,000
08	ZAMBIA	81.6%	18,243,000
09	MADAGASCAR	81.3%	27,101,000
10	PAPUA NEW GUINEA	81.2%	8,851,000

48 **SOURCES:** MANOCHI ANALYSIS; ITU; GSMA INTELLIGENCE; EUROSTAT; GOOGLE'S ADVERTISING RESOURCES; CNNIC; LOCAL GOVERNMENT AUTHORITIES; UNITED NATIONS. **NOTES:** FIGURES IN THE "% OF POP. OFFLINE" COLUMN REPRESENT THE PERCENTAGE OF THE POPULATION THAT DOES NOT YET USE THE INTERNET. ABSOLUTE VALUES HAVE BEEN ROUNDED TO THE NEAREST THOUSAND. (1) THE INTERNET (AT LEAST AS THE REST OF THE WORLD KNOWS IT) REMAINS BLOCKED FOR EVERYDAY CITIZENS IN NORTH KOREA. **COMPARABILITY:** SOURCE AND BASE CHANGES. ALL FIGURES USE THE LATEST AVAILABLE DATA, BUT SOME SOURCES DO NOT PUBLISH REGULAR UPDATES, SO FIGURES MAY UNDER-REPRESENT ACTUAL USE. SEE NOTES ON DATA.

we are social MANOCHI

datareportal.com/reports/digital-2026-six-billion-internet-users

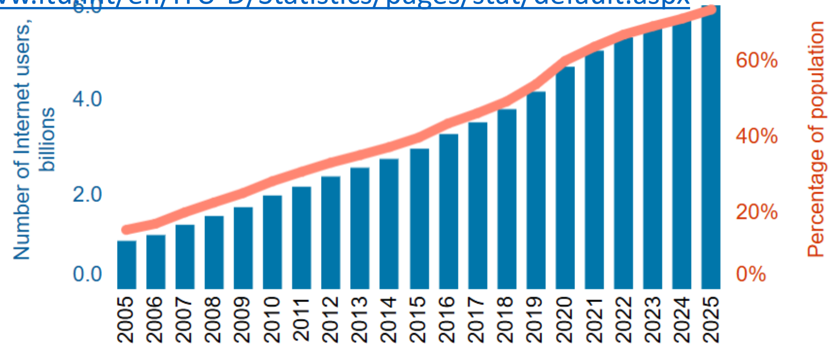
DATAREPORTAL

DIGITAL 2026: INTERNET USERS PASS THE 6 BILLION MARK

15 OCTOBER 2025 · SIMON KEMP

Individuals using the Internet

SOURCE: https://www.itu.int/en/ITU-D/Statistics/pages/stat/default.aspx



Source: https://datareportal.com/reports/digital-2026-mid-year-global-update-report

PROPOSED SOLUTION

GOVERNMENTS AND TELECOMMUNICATION COMPANIES STRATEGIES

Why Telecommunication Companies are Strategic

- **Assets already in place:**
  - 300,000+ tower sites (real estate, security, existing power backup)
  - Subsea cable landings and metro fibre rings
  - Mobile Money platforms (potential for distributed data aggregation)
- **Unique role:** Telcos can host **edge AI compute** at the last mile, the closest point to users and radio signals.
- **Scientific Publication:** Marinova and Leon-Garcia (2024)
  - Open RAN is revolutionizing the telecom space by introducing a framework based on virtualization and openness.
  - The network is optimized using RAN intelligent controllers (RICs), which can take data-driven, closed-loop actions in a RAN built in a multi-vendor, interoperable environment."
- ***CONCLUSION:*** ***No other industry has this physical footprint. Cloud providers do not own towers. Telcos are the missing middle.***

Why Governments Are Strategic

- **Policy levers:**
  - Spectrum licensing (can mandate AI-readiness)
  - Universal Service Funds (billions of dollars, currently underutilised)
  - Import duties (can waive on AI hardware)
  - Cross-border data and roaming regulations
- **Unique role:** Only governments can **de-risk private investment** and create continent wide standards.
- **Scientific Publication:** Gwagwa, Kazim and Kachali (2023)
  - African governments need to create policies and frameworks which create an enabling environment for the development and deployment of AI in Africa.
  - This includes creating frameworks and guidelines that do not exacerbate inequalities.
- ***CONCLUSION:*** ***Telecommunication Companies cannot fix national power policy or waive their own taxes. That is government work and it is essential.***

PROPOSED SOLUTION

THE RAN INTELLIGENT CONTROLLER (RIC) ARCHITECTURE

- **What is the RIC?**
  - A software platform that resides at or near the base station, using AI to make real-time decisions about radio resources without waiting for a distant cloud.

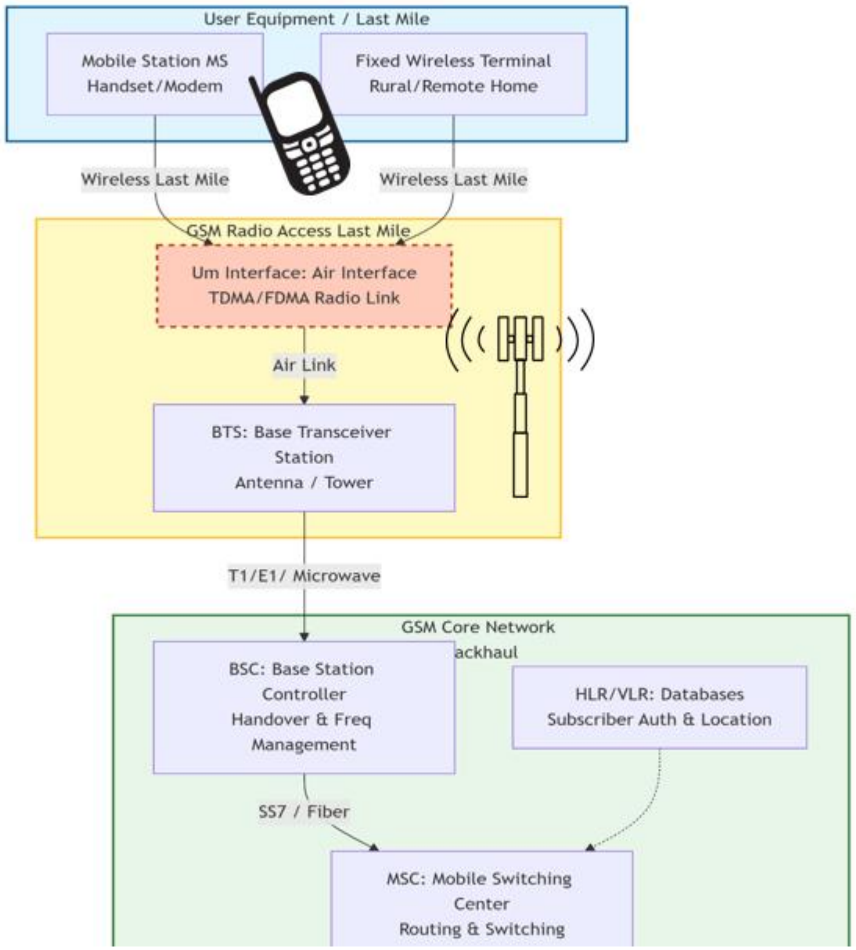
RIC Type	Timescale	Function	O-RAN Component
Near-RT RIC	10ms–1s	Handover, interference mgmt, dynamic spectrum sharing	xApps
Non-RT RIC	>1s (minutes–hours)	Policy, model training, long-term optimization	rApps

- **Scientific Publication:** Marinova and Leon-Garcia (2024)
  - They focus on the intelligence aspects by providing an in-depth view of the near-RT and non-RT RICs specified by the O-RAN Alliance, including the architecture and interfaces.
  - The argue that AI/ML approaches and frameworks are useful in the O-RAN context for applications (xApps and rApps) that programmatically and autonomously control and optimize the network through the RICs.
- ***CONCLUSION:*** ***The RIC is the autopilot for the radio network. In Africa, where backhaul and power are unreliable, this local intelligence is not optional but it is survival.***

PROPOSED SOLUTION

Distributed Artificial Intelligence Control Plane - Architectural Consideration

- **Centralised cloud model (fails in Africa)**
  - All telemetry to one data centre
  - High latency, backhaul congestion, single point of failure.
- **Distributed three-tier model (recommended)**
  - **Tier 1 – Edge (RIC at tower):** Real-time, <10ms. Handover, power sleep, interference mitigation.
  - **Tier 2 – Regional (metro area):** Near-real-time, seconds–minutes. Traffic prediction, model updates.
  - **Tier 3 – Central (national cloud):** Hours. Heavy model training, long-term analytics.
- **Scientific Publication:** Shojaei, Dán and Kóczy (2024)
  - Decentralised baseband functions across multiple locations require efficient placement of baseband and user plane functions in conjunction with Multi-Access Edge Computing (MEC) to accommodate diverse 5G services.
- ***CONCLUSION: Do not build one expensive central AI cloud. Build thousands of cheap edge AI brains on existing towers.***



PROPOSED SOLUTION

FEDERATED LEARNING FOR PRIVACY (Preserving Cross-Border AI)

- **What is Federated Learning?** A distributed machine learning framework where models are trained locally at each tower (or each country's operator) and only encrypted model parameters and never raw user data are exchanged to improve a global model.
- **African application:** Cross-border handover optimisation between Zambia (Airtel) and Zimbabwe (MTN) without violating data sovereignty laws.
- **Scientific Publication:** Javed, Mangues-Bafalluy, Zeydan and Blanco (2025)
    - This paper provides a robust and privacy-preserved federated learning (FL) framework in the open radio access network (O-RAN).
    - The solution prevents free-riding and issues with global model accuracy while preserving local data privacy and improving compliance in a multi-vendor environment.
- **Scientific Publication:** Nguyen, Phan and Vo (2025)
    - FGrINet can be implemented as an xApp architecture, aligning with O-RAN's goals of using distributed machine learning for intelligent, real-time RAN optimisation.
    - BSs can operate in favourable conditions to assist others, enhancing adaptability to diverse channel environments.
- ***CONCLUSION:*** ***Federated learning is not just a technical choice, it is a sovereignty requirement. Governments can trust this architecture because data never leaves the jurisdiction.***

OPPORTUNITIES (Peer Reviewed Scientific Publication) AI Enhancing Network Performance and Efficiency

Opportunity	Benefit	Key Reference
Predictive maintenance	Forecast tower battery/faults → 60% fewer truck rolls	Zhang, Li, Wu et al. (2023)
Dynamic network slicing with AI	64.2% cost reduction, 45.5% performance improvement	Xu, Zhang, Shen, Du (2025)
Energy-aware AI / sleep control	Predict traffic → idle base station sleep → 30% less diesel	López-Pérez et al. (2023)
Federated learning for RAN	Privacy-preserved cross-border optimization	Javed et al. (2025); Nguyen et al. (2025)
Deep RL for slice enforcement	18.5% throughput improvement with QoS guarantees	Shen, Zhang, Zhao, Gong (2022)
Autonomous closed-loop automation	35% OPEX reduction, 42% efficiency improvement	Rallanandi (2021)

OPPORTUNITIES (Peer Reviewed Scientific Publication) AI Enhancing Network Performance and Efficiency

- **Predictive maintenance:** Zhang, Li, Wu, Jiao, Ren, Hu and Liu (2023), *Entropy*, Vol. 25(6)
  - Proposed a layered method to predict hardware component failures using alarm logs, detecting faulty base stations and components with high precision.
  - Predictive maintenance is important from both a cost reduction and a network performance perspective.
- **Network slicing (AdaSlicing):** Xu, Zhang, Shen and Du (2025), *IEEE INFOCOM*
  - Propose AdaSlicing, an adaptive network slicing system that substantially outperforms state of the art works with **64.2% cost reduction and 45.5% normalized performance improvement**.
- **Energy efficiency (SRCON):** López-Pérez, De Domenico, Piovesan and Debbah (2023)
  - The SRCON framework harnesses live network data and employs ML and expert based models to accurately predict network energy efficiency for any energy carrier shutdown configuration, eliminating reliance on expensive expert knowledge
- **Deep reinforcement learning for network slicing:** Shen, Zhang, Zhao and Gong (2022), *IEEE JSAC*, Vol. 40(8)
  - LinkSlice was designed as an iterative two-stage DRL-based algorithm that improves throughput by 18.5% while ensuring link-level soft slice isolation and guaranteeing diverse QoS requirements.
- **Closed-loop automation in private 5G:** Rallanandi (2021):
  - A framework for closed-loop automation reduces operational expenditure by **35% while improving network efficiency by 42%** compared to traditional management approaches.
- ***CONCLUSION: Every single item saves money. AI is not a cost, it is an operational efficiency tool with peer reviewed, quantified returns.***

CHALLENGES REQUIRING JOINT GOVERNMENT AND TELECOMMUNICATIONS COMPANIES ACTION

Challenge	Who Must Act	Specific Action
High import tariffs on AI hardware	Government	Waive duties on edge inference modules, GPUs, batteries
USFs exclude edge compute	Government	Revise Universal Service Fund rules to fund compute, not just fibre
Unreliable grid power	Both	Co-invest in solar + battery storage at tower sites
Lack of MLOps & telecom AI skills	Both	Fund university courses, vocational training
Fragmented cross-border data rules	Government (AU, RECs)	Harmonise data protection and roaming for federated learning
High backhaul cost	Telcos + Government	Infrastructure sharing mandates; open access to state-owned fibre
Vendor lock-in (proprietary RAN)	Both	Mandate open RAN with AI interfaces in public tenders

- **Scientific Publication:** Ansarullah, Saif, Andrabi, Kumhar, Kirmani and Kumar (2022)
  - 5G and beyond wireless systems must address unpredictable and seemingly unsolvable challenges involving massive volumes of data, requiring AI-enabled system design for resource usage, protection, versatility and expandability.
- **CONCLUSION:** ***No single actor solves any of these alone. The partnership is not optional, these challenges are technically solvable but require coordinated policy and investment.***

RECOMMENDATIONS FOR AFRICAN GOVERNMENTS

- **Classify edge AI nodes as critical digital infrastructure** – fast-track permits, priority grid connection, tax holidays.
- **Revise Universal Service Funds** – allow spending on compute, energy storage, and AI modules. Require AI-readiness in all subsidised infrastructure.
- **Remove import duties** on AI inference hardware (e.g., Google Coral, NVIDIA Jetson) and lithium batteries.
- **Create national AI network sandboxes** – regulatory safe zones for testing federated learning, network slicing, and cross-border AI.
- **Mandate open RAN with AI interfaces** in all state-owned or state-subsidised network procurement.
- **Support regional data trust frameworks** (e.g., African Union) to enable legal, secure cross-border federated learning.
- **Scientific Publication:** Gwagwa, Kazim and Kachali (2023)
    - Regulatory sandboxes could ensure the protection of rights without stifling innovation.
    - Investment into research and development is needed to enable African procurement of domestic and regionally developed AI tools rather than merely consuming technology from the Global North.
- **CONCLUSION:** *Policy is the cheapest infrastructure. A tax waiver costs nothing but unlocks billions in private investment.*

RECOMMENDATIONS FOR TELECOMMUNICATIONS COMPANIES

- **Deploy low-power AI inference at tower level** – start with top 20% of sites (highest traffic, most power instability).
- **Adopt open RAN architectures** – avoid vendor lock-in; allow third-party AI xApps (decision trees and other ML models have proven effective for network slicing optimisation).
- **Share infrastructure** (towers, fibre, power) with competitors to lower edge compute deployment costs.
- **Upskill network engineers** – train in time-series ML, anomaly detection, MLOps, and federated learning.
- **Launch "AI-as-a-Service"** – sell inference capacity from edge nodes to local startups, agritech, healthtech.
- **Participate actively in government-led sandboxes** – co-develop cross-border AI optimisation models.
- **Scientific Publication:** Zakariyya, Hussein and Al-Rimy (2024)
  - The integration between AI and OpenRAN holds great promise for intelligent resource allocation within 5G networks.
  - The Decision Tree model displays powerful classification abilities by effectively and efficiently classifying network slices to enhance resource management processes.
- **CONCLUSION:** *Open RAN plus AI is not just cheaper but it is smarter. Telecommunications Companies that move first will gain a decisive competitive advantage.*

EXAMPLES OF ZAMBIA AI USE CASES FOR THE SANDBOX

- AI-Driven Biometric Cattle Identification and Traceability System for Smallholder Farmers in Zambia
    - Simbalashe Kapundu and Mutinta Chijoka – MSc
    - Sponsored by NSTC and African Centre for Technology Studies (ACTS)
- Using IoT and Machine Learning for Predictive Maintenance, Monitoring and Preventing Vandalism of Parastatal Distribution Infrastructure
    - Chimuka Moonde - PhD **Candidate**
    - Ministry of Science and Technology Sponsorship
- Large Language Model for ZNPHI Disease Surveillance – Measles Immunity Study
    - Mulima Chibuye - PhD **Graduand**
    - **ZNPHI Sponsored Project**

African Use Cases

- **Rural tower energy AI** – Solar/battery site uses RIC to throttle non-critical processing and sleep idle radios → 30% less diesel.
- **Cross-border handover optimisation** – Federated learning between Kenya and Uganda improves handover success from 92% to >99%.
- **Mobile Money transaction routing** – Real-time network slicing cuts transaction time from 15s to 2s.

USE CASE 1: (1/2)

AI-Driven Biometric Cattle Identification and Traceability System for Smallholder Farmers in Zambia

Problem Overview

- Smallholder farmers dominate cattle farming in Zambia
- Livestock theft causes major economic losses
- Traditional methods (ear tags, branding) are:
  - Easily removed or tampered with
  - Unreliable for long-term tracking
- Poor record-keeping limits traceability and financial access
- Existing tech (RFID) is costly and not practical in rural areas

AI-Based Biometric Solution

- AI-driven cattle identification using **facial recognition**
- Uses **Convolutional Neural Networks (CNNs – ResNet-50)**
- Mobile application for real-time cattle identification
- Offline-first design for rural environments
- Cloud integration for storage and traceability

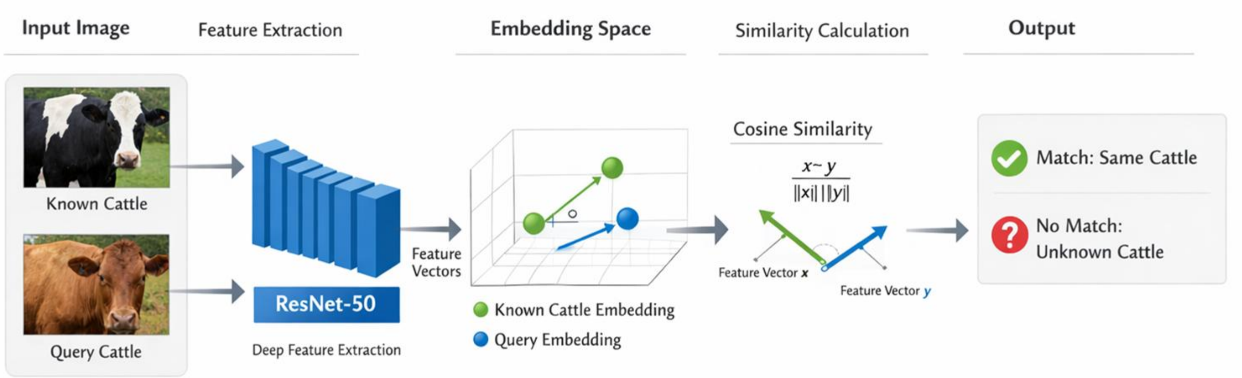


Figure: 1 cattle Identification pipeline

"Identify cattle using their natural features — not removable tags"

USE CASE 1: (2/2)

AI-Driven Biometric Cattle Identification and Traceability System for Smallholder Farmers in Zambia

AI Methodology & System Pipeline

- **Model:** ResNet-50 (Transfer Learning)
- **Input:** Cattle facial images (real-world dataset)
- **Preprocessing:(Roboflow)**
  - Normalization
  - Data augmentation (lighting, rotation, noise)
- **Feature Extraction:** Deep embeddings from CNN
- **Similarity Matching:** Cosine similarity
- **Decision Logic:**
  - High similarity → Known cattle
  - Low similarity → Unknown (open-set recognition)

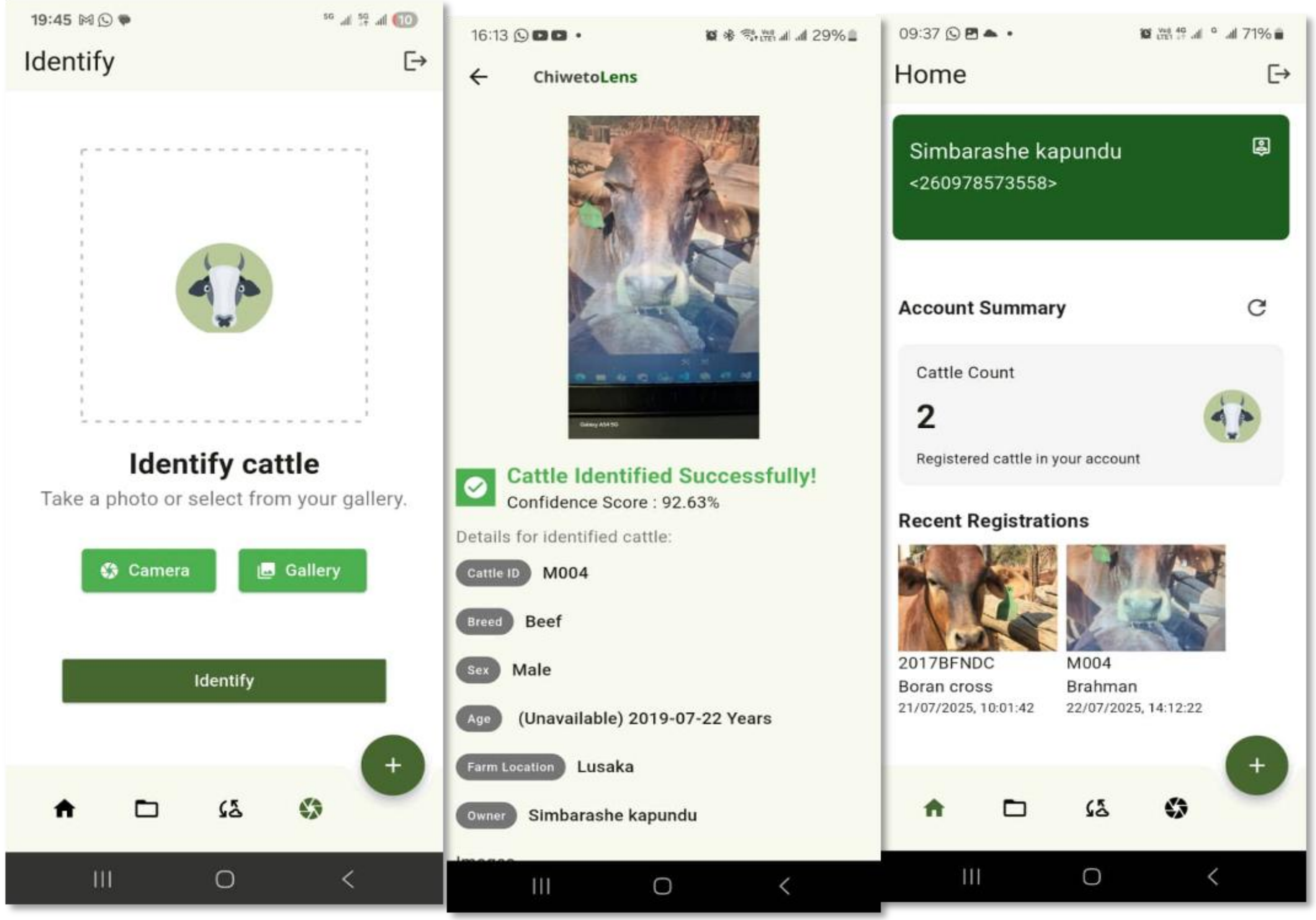


Figure 3: Mobile application prototype

USE CASE 2: (1/2)

Using IoT and Machine Learning for Predictive Maintenance, Monitoring and Preventing Vandalism of Parastatal Distribution Infrastructure

Problem Statement

Critical energy and cell tower infrastructure across Africa faces persistent threats that compromise reliability and safety. Current monitoring approaches are inadequate for addressing these challenges effectively.

Vandalism Crisis

- Frequent theft and vandalism of transformers, copper cables, and distribution equipment causing widespread service disruption

Reactive Monitoring

- Traditional surveillance systems respond after incidents occur rather than preventing them in real-time

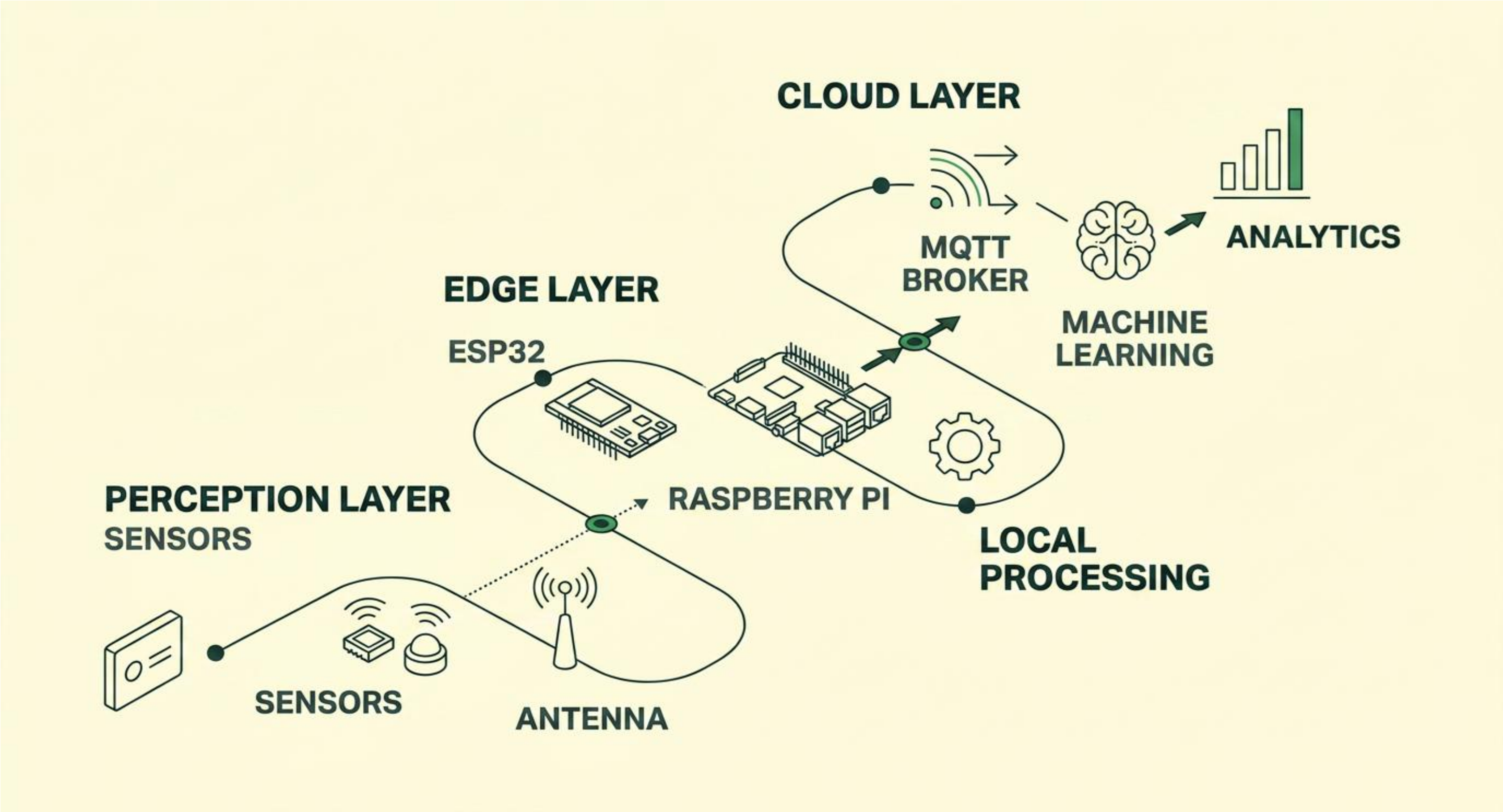
Financial Impact

- Extended outage durations resulting in significant revenue loss and costly equipment replacement

Intelligence Gap

- Absence of autonomous, intelligent systems capable of adaptive threat detection and response

System Architecture & Deployment



The **Perception Layer** consists of an array of intelligent sensors for data collection.

Data is then fed to the **Edge Layer**, powered by ESP32/Raspberry Pi devices, where initial processing and rule-based AI analysis occur.

Finally, data is securely transmitted via MQTT over LoRA/LTE to the **Cloud Layer** for advanced analytics, machine learning (Isolation Forest/Random Forest), and dashboard visualization, enabling continuous learning and adaptive threat detection.

Deployment & AI Strategy

Flexible Deployment

Deployed across **urban, peri-urban, and rural sites**, ensuring broad coverage regardless of geographical challenges.

Resilient Hardware

All hardware is **solar-powered** and housed in **IP65 weatherproof enclosures** for durability in harsh conditions

Hybrid AI

Edge AI provides immediate rule-based threat detection, while **Cloud AI** uses Isolation Forest/Random Forest for deeper anomaly analysis and **Continuous Learning**.

USE CASE 3: (1/2)
Large Language Model for ZNPHI Disease Surveillance – Measles Immunity Study

THE PROBLEM:

- IDSR guidelines exist in hundred-page PDFs—not portable or searchable

- HCWs cannot rapidly access case definitions under clinical load

- **2023-24 cholera outbreak:** 17,000+ cases, fatality rates above WHO threshold

- **Time-to-Answer:** Minutes to Hours

THE SOLUTION:

ZNPHI AI Surveillance Assistant

A RAG chatbot delivering point-of-care, guideline-grounded answers

✓ **Multi-Disease:** Measles (deployed), Cholera, TB, Malaria, HIV/AIDS
✓ **Multilingual:** Nyanja & Bemba (66% of Zambians)
✓ **Verifiable:** 100% of claims cite approved ZNPHI sources

▶ Time-to-Answer: <30 Seconds

USE CASE 3: (2/2)

Large Language Model for ZNPHI Disease Surveillance – Measles Immunity Study

Technical Architecture & Implementation

Production Stack:

- Google Vertex AI (Gemini 2.5 Flash)
- text-embedding-004 embeddings
- FAISS for vector search
- LangChain for orchestration
- FastAPI on Google Cloud Run
- BigQuery logging for observability

Key Innovations:

✓ Disease-agnostic indexing
✓ Intelligent disease router
✓ Hybrid retrieval (BM25 + dense)
✓ Structured citations
✓ Clinical safety & PII scrubbing

12-Month Roadmap: 5 Phases | P0-1 Foundation | P2 Cholera | P3 Malaria & TB | P4 HIV/AIDS | P5 Multilingual Pilot

REFERENCES

- S. Marinova and A. Leon-Garcia, “Intelligent O-RAN Beyond 5G: Architecture, Use Cases, Challenges, and Opportunities,” IEEE Access, vol. 12, pp. 27088–27114, 2024.
- M. Xu, D. Zhang, X. S. Shen and Y. Du, “AdaSlicing: Adaptive Online Network Slicing Under Continual Network Dynamics in Open Radio Access Networks,” in Proc. IEEE INFOCOM, London, UK, 2025.
- I. S. Ansarullah, S. M. Saif, S. A. Andrabi, S. H. Kumhar, M. M. Kirmani and D. P. Kumar, “A Roadmap for AI Enabled System Design for 5G, 6G, And Beyond Communication: Issues And Challenges,” in Proc. ICC, 2023.
- D. López Pérez, A. De Domenico, N. Piovesan and M. Debbah, “Data driven Energy Efficiency Modelling in Large scale Networks: An Expert Knowledge and ML based Approach,” arXiv preprint arXiv:2401.00443, 2023 (submitted to IEEE TMLCN).
- F. Javed, J. Mangues Bafalluy, E. Zeydan and L. Blanco, “Trustworthy Reputation for Federated Learning in O RAN Using Blockchain and Smart Contracts,” IEEE Open Journal of the Communications Society, vol. 6, pp. 1343–1362, 2025.
- B. Rallanandi, “Autonomous Private 5G Networks For Industry 4.0: AI Native Operations And Closed Loop Automation,” Journal of International Crisis and Risk Communication Research, pp. 311–326, 2021.
- Y. Shen, Y. Zhang, J. Zhao and Y. Gong, “LinkSlice: Fine Grained Network Slice Enforcement Based on Deep Reinforcement Learning,” IEEE Journal on Selected Areas in Communications, vol. 40, no. 8, pp. 2378–2394, Aug. 2022.
- K. Zhang, Y. Li, D. Wu, Y. Jiao, X. Ren, B. Hu and S. Liu, “Predicting Network Hardware Faults through Layered Treatment of Alarms Logs,” Entropy, vol. 25, no. 6, p. 917, Jun. 2023.
- D. Das, A. Farid and S. Saha, “Connectivity for AI enabled cities – A field survey based study of emerging economies,” arXiv preprint arXiv:2501.09479, 2025.
- A. Gwagwa, E. Kazim and E. Kachali, “Harnessing artificial intelligence for inclusive and sustainable economic growth in Africa: A path towards Agenda 2063,” Research ICT Africa / AUDA NEPAD, 2023 (peer reviewed policy research).
- T. T. Nguyen, T. V. Phan and N. S. Vo, “Decentralized Federated Learning for GNN Based Channel Estimation With DM RS in O RAN,” in Proc. IEEE ICC, Barcelona, Spain, 2025.
- S. Zakariyya, H. Hussein and R. Al Rimy, “Evaluation of AI Models for Intelligent Network Slicing in OpenRAN: A Comprehensive Benchmarking Approach,” in Proc. ICECE, Bangkok, Thailand, 2024.

Thank you

Questions