STRENGTHENING KENYA'S DATA ECOSYSTEM FOR AI

MSME and Public Service Delivery Use Cases

GSMA Mobile for Development

M4D

GSMA

The GSMA is a global organisation unifying the mobile ecosystem to discover, develop and deliver innovation foundational to positive business environments and societal change. Our vision is to unlock the full power of connectivity so that people, industry, and society thrive. Representing mobile operators and organisations across the mobile ecosystem and adjacent industries challenges, the GSMA delivers for its members across three broad pillars: Connectivity for Good, Industry Services and Solutions, and Outreach. This activity includes advancing policy, tackling today's biggest societal issues, underpinning the technology and interoperability that make mobile work, and providing the world's largest platform to convene the mobile ecosystem at the MWC and M360 series of events.

We invite you to find out more at www.gsma.com

UK International Development

Partnership | Progress | Prosperity

This material has been funded by UK International Development from the UK Government and is supported by the GSMA and its members. The views expressed do not necessarily reflect the UK Government's official policies.

GSMA Central Insights Unit

The Central Insights Unit (CIU) sits at the core of GSMA Mobile for Development (M4D) and produces in-depth research on the role and impact of mobile and digital technologies in advancing sustainable and inclusive development. The CIU engages with public and private sector practitioners to generate unique insights and analysis on emerging innovations in technology for development. Through our insights, we support international donors to build expertise and capacity as they seek to implement digitisation initiatives in low- and middle-income countries through partnerships within the digital ecosystem.

Contact us by email: centralinsights@gsma.com

Author
Zarah Udwadia, GSMA Mobile for Development

Contributors
Az Choudhary, Mobile for Development
Jamie Robertson, GSMA Mobile for Development
Nigham Shahid, GSMA Mobile for Development
Ruth Orbach, GSMA Mobile for Development

Acknowledgements
This report draws on research conducted for the GSMA by Cynthia Amol, Dr. Lilian Wanzare and Dr. Savita Bailur.

We would also like to thank Wilfred Oluoch from the FCDO Kenya for his support.

We are grateful to all the individuals and organisations who shared their insights and expertise during this research. A list of those consulted can be found at the end of the report.

CONTENTS

List of tables

List of figures

List of acronyms

AI	Artificial Intelligence
API	Application Programming Interface
CHP	Community Health Promotor
DPA	Data Protection Act
DPI	Digital Public Infrastructure
GIS	Geographic Information System
GPS	Global Positioning System
ICT	Information and Communication Technology
IoT	Internet of Things
KODI	Kenya Open Data Initiative
KRA	Kenya Revenue Authority
KNBS	Kenya National Bureau of Statistics
ML	Machine Learning
MNO	Mobile Network Operator
MSME	Micro, Small and Medium-sized Enterprise
NLP	Natural Language Processing
ODPC	Office of the Data Protection Commissioner
SLM	Small Language Model
SME	Small and Medium-sized Enterprise

Definitions

Algorithm	A process or set of rules to be followed in calculations, especially by a computer, to solve a problem.[1]
Application programming interface (API)	A set of rules or protocols that enables software applications to communicate with each other to exchange data, features and functionality.[2]
Artificial Intelligence	Artificial intelligence (AI) is a set of technologies that enables computers and machines to simulate learning, comprehension, problem solving, decision making, creativity and autonomy.[3]
Data	Data is a collection of raw numbers, words, observations, measurements, or other information, collected from various sources and in various formats. Data can be structured (organised in a clear, defined format), unstructured (lacking a strictly defined format) or semi-structured. A dataset is an organised collection of data, for retrieval, analysis or use.[4]
Digital Public Infrastructure	Foundational, re-usable digital building blocks, such as digital payments, ID and data sharing, designed for the public benefit.[5]
Generative AI (GenAI)	A type of AI that involves generating new data or content, including text, images or videos, based on user prompts and by learning from existing data patterns.[6]
Interoperability	Interoperability enables different information technology (IT) systems to exchange data with minimal end user intervention, made possible by using common standards that define how data is formatted and exchanged.[7]
Natural language processing (NLP)	A field of machine learning in which machines process natural language as spoken and written by humans, instead of the data and numbers normally used to program computers.[8]
Synthetic data	Artificial data designed to mimic real-world data, generated through statistical methods or AI techniques like deep learning and GenAI.[9]

1 Definition by GSMA.
2 Definition by IBM.
3 Definition derived from IBM.
4 Definition by IBM.
5 Definition by Gates Foundation
6 Definition by GSMA.
7 Definition by IBM.
8 Definition by GSMA.
9 Definition by IBM.

EXECUTIVE SUMMARY

Kenya's AI ambitions depend on strong data foundations

Kenya has a vibrant digital landscape with a thriving homegrown tech industry, active global Big Tech companies and a strong digital economy. The country also stands out on the continent for its early development of AI-based solutions and AI infrastructure. Kenya's National AI Strategy (2025–2030) states ambitions to be a regional AI leader and to use AI to support inclusive socio-economic development in eight priority sectors.

The strategy recognises **data** as a key pillar and building block of AI systems, determining their performance, generalisability, reliability and impact, particularly in high-stakes settings.

For Kenya to deliver on its AI ambitions, it needs a strong data ecosystem that goes beyond data and datasets to include data systems, governance, infrastructure and the individuals and institutions involved. While the capabilities and demands of emerging technologies have helped expand local datasets such as language and geospatial data, gaps remain. Without hyperlocal data, the full diversity of the country cannot be captured and AI solutions for vulnerable populations risk being discriminatory and ineffective. This is particularly important for **public service delivery** and **micro-, small and medium enterprises (MSMEs)**, both national priorities and a focus of this study. By mapping the data ecosystem in these two sectors, this research aims to understand the data-related barriers and opportunities to developing impactful AI solutions for citizens, government and businesses.

AI readiness requires a stronger, more inclusive data ecosystem

The statistical and administrative data held by national and county governments are the most significant sources for public service delivery, but there are obstacles to using it effectively for AI applications. Data held by public sector institutions typically needs extensive cleaning, labelling, annotation, structuring, standardisation and sometimes even digitisation. Data standards are not consistently adopted, leaving the public sector data ecosystem without clear quality assurance mechanisms and evaluation processes. Innovative public service delivery in Kenya currently depends on combining datasets from different sectors. Yet, despite progress, digital data systems are fragmented and not interoperable, making it difficult to link datasets, establish automated data-sharing mechanisms between institutions and share up-to-date data.

Non-profit and international development partners play a key role in public service delivery, funding data partnerships, tools and data collection efforts.

Private sector data is critical, but partnership incentives need to be clearer

The private sector, however, is the main driver of AI development in Kenya. For example, mobile network operators (MNOs) and social media platforms hold vast amounts of granular, digital, high-quality data that can be valuable for both public service delivery and MSMEs. They also provide the necessary infrastructure and skills training. However, to facilitate effective data sharing and mutually beneficial partnerships, incentives need to be aligned and sustainable funding mechanisms need to be in place.

MSME data is growing, but informality limits visibility and representation

MSME AI applications rely on a mix of public-sector data on business administration and operations, MSMEs' own transaction and administrative data, as well as private-sector data, including from mobile money platforms, e-commerce and financial institutions. However, high levels of informality and low digitisation limit both the availability and representativeness of these datasets. Stronger digital data collection processes are needed, along with support for MSMEs to expand their use of digital tools. This not only involves building capacity and skills but also trust, new behaviours and driving systemic change and incentives.

Skills, infrastructure and trust remain cross-cutting barriers

Across both sectors, fundamental gaps in skills (like data collection and analysis) and infrastructure (connectivity) are limiting the adoption of stronger data practices. At the same time, safe and responsible data collection depends on the trust of data owners and subjects. As AI use cases demand increasing amounts of data, it is important to ensure they trust the data collection process, how their data is ultimately used and are valued fairly for it. Strong governance and new African-centric licensing models also play a role.

AI use cases show some potential, but depend on stronger shared data foundations

In the public sector, AI-driven solutions can help improve the efficiency, accountability and transparency of service delivery. However, the risks are high, tolerance for error and experimentation is justifiably low and reliance on public sector data makes it difficult for the private sector and startups to innovate. Use cases like AI-enabled support for community health worker triage demonstrate the importance of interoperable data systems and ID frameworks, as well as the buy-in and trust of frontline workers, as both data generators and end users of AI systems.

In the MSME sector, Kenya's vibrant fintech innovation ecosystem is backing AI innovation, with use cases like alternative credit scoring, supply chain management and advisory support. However, informal, resource-strapped MSME owners generate the lowest amounts of digital data, leading to risks of bias when using AI models trained on such data.

The time and resources required to make data AI-ready can deter innovators and researchers from even attempting to access government data, especially at the county level. This leaves vital datasets largely untapped, which are especially valuable in devolved sectors like MSMEs, health and agriculture. Innovators often end up building their own contextually relevant data pipelines, a process that can be operationally and financially taxing. Because data is treated as a competitive advantage, data collection and digitalisation efforts are often duplicative rather than collaborative. Ultimately, collaboration across actors, sectors and the data value chain is needed to build a strong data ecosystem capable of supporting impactful and equitable AI solutions.

While governments, communities, MSMEs and citizens all benefit from data, these benefits are often not apparent across the data value chain, and data often does not garner the same attention as other, more visible parts of the AI ecosystem. Willingness to invest in data is low in resource-constrained settings, but this is the context in which data-driven decision making can add the most value. Successful initiatives to strengthen the data ecosystem have come from demonstrating real-world value: an AI-solution pilot developed for a Ministry despite the limited data available, or a county government seeing and learning from the successful implementation of AI and data systems in another county.

1. INTRODUCTION

Kenya's digital landscape

Kenya stands out as one of Africa's leading technology hubs, underpinned by almost universal mobile broadband coverage.[10] More than 90% of men and women own a mobile device, and smartphone penetration, currently at 50% of the population, is expected to rise to 81% by 2030.[11] Kenya's advanced digital landscape has paved the way for a rich homegrown technology ecosystem with pioneering mobile-first innovations, particularly in the digital financial services (DFS) sector. Today, the information and communication sector attracts the highest share of foreign direct investment (FDI)[12], while venture capital (VC) raised by Kenyan startups in 2024 represented more than a third of Eastern Africa's total VC investment.[13] The country's digital economy continues to grow, projected to contribute KSH 662 billion to GDP by 2028.[14]

>90%
Men and women who own a mobile device

50%
Smartphone penetration in Kenya

81%
Smartphone penetration expected by 2030

10 Ministry of ICT and the Digital Economy. (2023). Kenya Artificial Intelligence Strategy 2025–2030.
11 GSMA. (2025). The Mobile Gender Gap Report.
12 Kenya National Bureau of Statistics. (2024). 2024 Foreign Investment Survey Report.
13 Ashiru, G. (2025). "Kenya's Tech Sector Becomes Top Destination for Foreign Direct Investment". Tech in Africa.
14 GSMA. (22 October 2024). "Kenya's Digital Economy to Contribute KSH 662 Billion to GDP by 2028, Driven by Policy Reforms".

Kenya's AI ecosystem

Building on this strong digital foundation, Kenya is already putting AI to use to drive economic growth and address development challenges, particularly in agriculture, food security and energy.[15] Kenya has one of the highest concentrations of AI startups on the continent,[16] with innovators developing AI-powered solutions tailored to their local contexts.[17] In addition to the integration of AI in the public and private sector, Kenyans are integrating AI in their daily lives, with more than a quarter reportedly using ChatGPT daily.[18]

Kenya ranks sixth among Sub-Saharan African countries according to Oxford Insights' 2024 Government AI Readiness Index, but when compared to countries around the globe, its ranking drops to 93 out of 188, scoring lower than North American and Western European countries in particular.[19] Local AI infrastructure remains a gap, reducing the ability of innovators and researchers to scale AI development. Without adequate supercomputers (HPC) and graphic processing units (GPUs), the AI ecosystem relies largely on foreign cloud providers, resulting in lower latency. However, the government, often in partnership with the private sector, is expanding AI infrastructure,[20] from the development of data centres to edge computing,[21] broadband, GPU capacity and local semi-conductor manufacturing.[22]

Another critical barrier is uneven digital access. Women face a 16% gender gap in smartphone ownership and a 22% gender gap in mobile internet adoption.[23] Rural populations continue to lag in mobile internet adoption, although this rural-urban gap has been declining significantly year on year.[24] With the rapid growth of AI, there is a risk that women and rural populations will be left even further behind, widening the digital divide. Their smaller digital footprints also mean they are missing from the datasets used to train AI models, resulting in AI systems that do not represent their realities.[25]

To shape the country's AI trajectory, the Ministry of Information, Communications and the Digital Economy introduced the Kenya AI Strategy 2025–2030. The strategy lays out the country's ambitious vision of becoming a regional AI leader, with an emphasis on using AI for inclusive socio-economic development.[26] This vision is grounded in three pillars: Data; AI Digital Infrastructure; and AI Research and Development (R&D) and Innovation.

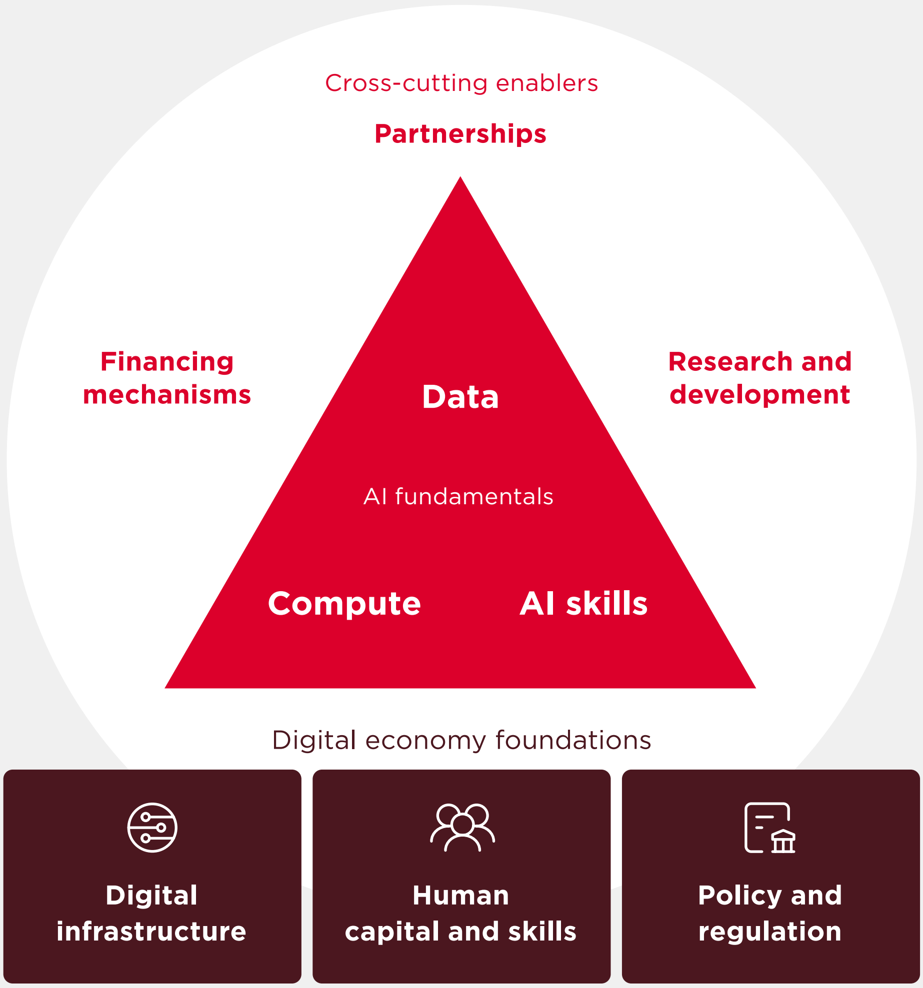
The importance of data

The fundamental building block of AI systems is high-quality text, image, audio, video or numerical data that is relevant to and representative of the context where it is being used. Deep learning models require particularly vast amounts of data for training and testing. Poor-quality data can have unseen, compounding negative effects on the safety, performance, reliability and cost of an AI model,[27] while datasets that are biased, incomplete or unrepresentative produce AI systems that reinforce existing inequalities.[28] As AI is increasingly applied to sectors like public service delivery, health and climate, AI-mediated predictions or decisions have major consequences on people's lives, and the stakes for data quality are higher than ever.[29] For instance, poor data practices have weakened the accuracy of models used to treat or predict diseases in the health sector, with the potential for real human harm.[30] In the agriculture sector, models trained on crop data that does not reflect local realities can lead to unsuitable predictions, impacting farmers' livelihoods.

15 GSMA. (2024). AI for Africa: Use cases delivering impact: Kenya Deep Dive.
16 Oloyede, J. (27 June 2025). "Agentic AI could be Africa's next big competitive advantage". TechCabal.
17 Ndemo, B. and Ragnet, M. (20 September 2025). "AI sovereignty in Kenya: Building a future rooted in local ownership and innovation". Brookings.
18 HAI Stanford. (2024). "Chapter 9: Public Opinion". Artificial Intelligence Index Report 2024.
19 Oxford Insights. (2024). Government AI Readiness Index 2024.
20 Adams, R. et al. (2025). AI in Africa: A Landscape Study. Global Center on AI Governance.
21 Ecofin Agency. (21 August 2025). "Kenya Invites Bids for Rural Broadband Rollout Under $390m Digital Economy Project".
22 U.S. Trade and Development Agency. (24 May 2024). "USTDA Partners with Kenya to Boost Semiconductor Manufacturing".
23 GSMA. (2025). The Mobile Gender Gap Report.
24 GSMA. (2025). Trends in Mobile Internet Connectivity: The State of Mobile Internet Connectivity 2025.
25 Stone, E. (8 February 2024). "AI shifts the goalposts of digital inclusion". JRF.
26 Ministry of ICT and the Digital Economy. (2023). Kenya Artificial Intelligence Strategy 2025–2030.
27 Sambasivan, N. et al. (2021). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. Conference on Human Factors in Computing Systems.
28 Global Partnership for Sustainable Development Data and GIZ. (2023). Artificial Intelligence Practitioners' Guide: Kenya.
29 Ibid.
30 Ibid.

Across Africa, historical and structural inequities have resulted in data poverty, or a dearth of high-quality digital data that adequately represents the continent's realities. Gaps in digital infrastructure, lower rates of digitalisation and digital access compared to high-income countries (HICs), and the historic marginalisation of Africa's languages compared to colonial languages,[31] have resulted in lower data generation, availability and access. For instance, there is 2,650 times more English content on the internet than African languages.[32] Even Kenyan public data is primarily in English despite Swahili also being a national language.[33] Growing recognition of the importance of locally generated data in building AI systems that reflect African languages, cultures and socio-economic realities has helped stimulate new data collection and curation initiatives, closing this gap.[34] Yet the hyperlocal language and domain data required for tailored applications are still lacking in some parts of Kenya.

Figure 1: The AI ecosystem framework



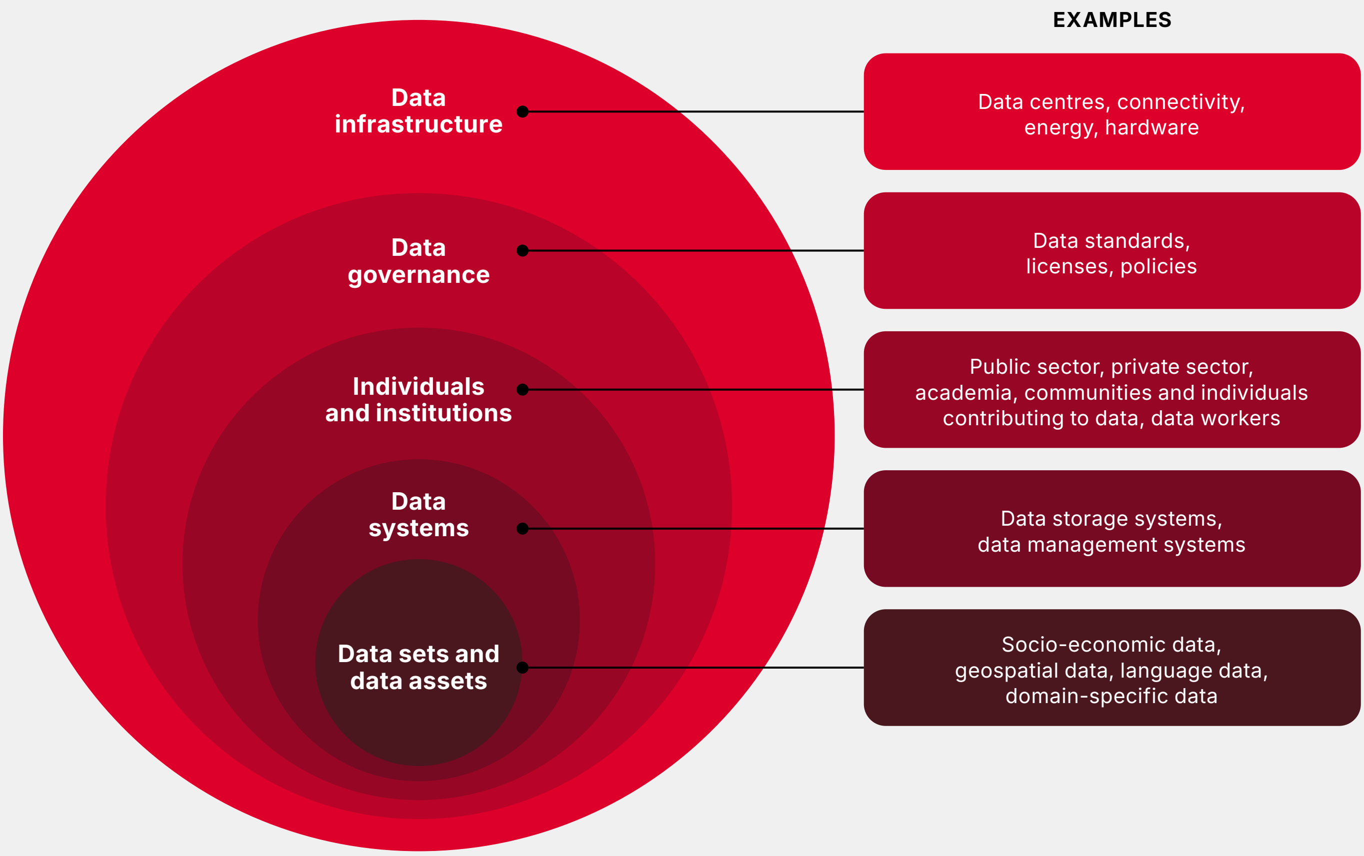
Source: GSMA AI for Africa

31 Marivate, V., Adebara, I. and Wanzare, L. (16 October 2025). "African languagees for AI: the project that's gathering a huge new dataset". DownToEarth.
32 Genesis. (2024). "AI in Africa: The state and needs of the ecosystem".
33 Global Data Barometer. (2025). 2nd Edition Report.
34 Strathmore University Centre for Intellectual Property and Information Technology Law (CIPIT). (2025). The State of AI in Africa Report.

While the rapid rise in AI deployment and adoption has created new data-related opportunities, challenges and needs, data-driven decision-making in policy and the public sector has long been recognised as essential for sustainable development. In 2013, the United Nations called for a "data revolution" to meet the world's complex global development needs[35] and, in the decade since, significant progress has been made in tracking the Sustainable Development Goals (SDGs), national and civil society data capacity, advanced use of technology and citizen-generated data, open data initiatives and coordinated global data standards. Recent continental frameworks have reiterated the importance of data for development and human rights, including UNESCO's 2024 Accra Statement on data[36] and the 2024 resolution of the African Commission on Human and Peoples' Rights (ACHPR).[37]

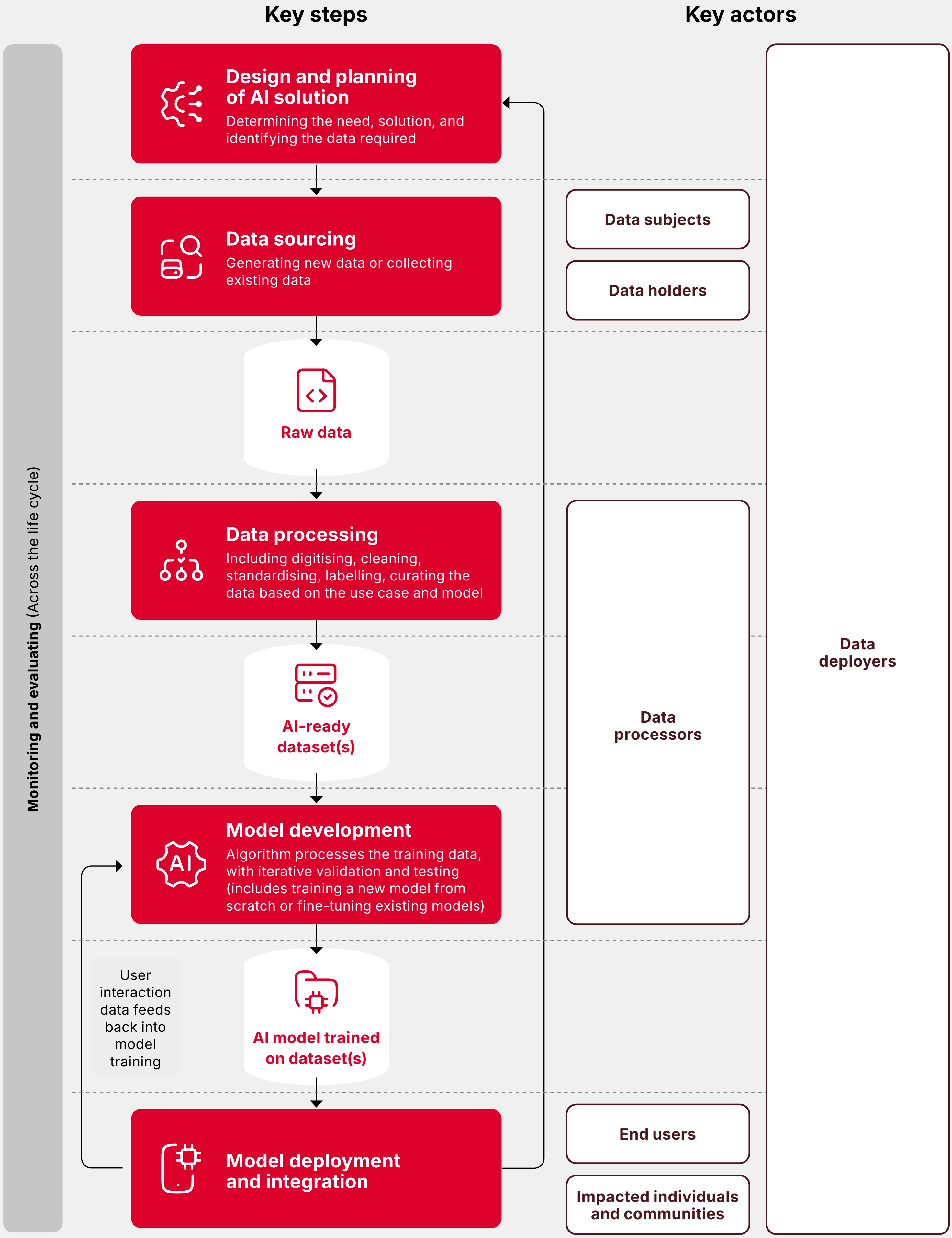
Data is also shaped by social, cultural, historical and political choices and contexts, all of which influence what gets counted, what does not, by whom and how.[38] This is particularly significant in the African context, where Western approaches to data collection and processing can disconnect African data from its history and meaning.[39] This study therefore looks at **data ecosystems**, which include both the technical and human elements involved in creating, managing, sharing, interpreting and using data (Figure 2).[40,41]

Figure 2: Elements of the data ecosystem



35 UN Economic and Social Council. (2014). Emerging issue: the data revolution. Report of the Secretary-General.
36 UNESCO. (2024). Harnessing the Power of Data: A Commitment to Strengthening Access to Information in the Digital Age.
37 The ACHPR is an organ of the African Union (AU). See: African Alliance for Access to Data. (18 November 2024). "Africa reaches a major milestone in its quest for data access".
38 Sambasivan, N. et al. (2021). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. Conference on Human Factors in Computing Systems.
39 Lemayian, D. and Gitau, S. (2025). Made in Africa: An African Perspective to the Design, Deployment and Governance of AI. Qubit Hub.
40 Toorajipour, R., Oghazi, P. and Palmié, M. (2024). "Data ecosystem business models: Value propositions and value capture with Artificial Intelligence of Things". International Journal of Information Management, 78.
41 Open Data Institute. (2019). Data Ecosystem Mapping Tool.

Figure 3: **Key stages of the AI and data life cycle**



Data for AI

While AI development often prioritises models and algorithms, the concept of data-centric AI[42] highlights the importance of the data used to train AI systems.[43] Training AI models introduces new demands to data quality, as models pick up false correlations and biases in the training data.[44] Data is critical across every stage of the AI life cycle,[45] as described in Figure 4.

Several frameworks provide guidance on data practices, such as the FAIR framework,[46] but AI calls for a degree of technical specificity these frameworks were not designed to have.[47] For instance, there is a lack of concrete, operational guidance on aspects like data annotation and bias mitigation. Also, factors that might be less important for traditional data quality, like label accuracy, bias and noise, can have significant impacts on the behaviour of AI models.[48] More recently, frameworks and tools such as the Open Data Institute (ODI)'s Framework for AI-Ready Data[49] and the World Bank's Data Quality and AI for Data, Data for AI programmes[50] make data discoverable, readable, accessible and usable by AI systems.[51] Dalberg Data Insights' People-Centred AI Playbook[52] demonstrates the importance of ensuring "people readiness" before deploying AI systems.

Figure 4: Data involved across the AI life cycle

Development

Training data
Processed specifically to train an AI model

Reference data
Augments training data with additional context or moderates the content of training data

Fine-tuning data
Used to adapt a pre-trained model to suit specialised use cases or contexts

Testing and validation data
Used to test a model while in development

Benchmarks
Datasets used to evaluate model performance

Data about the data used to develop models, or metadata
Includes information on the size, source and origins of datasets used

Deployment

User data
Data an AI model ingests and processes when deployed

Reference data
Retrieved and used to respond to queries rather than generating responses from the trained model alonea

Prompts
Inputs into an AI system to generate responses

Outputs from models
Data resulting from predictive or generative models

Monitoring

Data about models
Includes information on the technical attributes, performance metrics and ethical considerations of a model

Data about model usage and performance in context
Used to make improvements to models

Registers of model deployments
Lists of AI models used in different contexts (e.g. by a government)

Data about the AI ecosystem
Data about AI models, companies, risks and policies

Source: ODI, "A data for AI taxonomy".

42 van der Schaar Lab. (n.d.). What is Data-Centric AI?
43 Sambasivan, N. et al. (2021). "Everyone wants to do the model work, not the data work": Data Cascades in High-Stakes AI. Conference on Human Factors in Computing Systems.
44 Liang, W. et al. (2022). "Advances, challenges and opportunities in creating data for trustworthy AI". Nature Machine Intelligence, 4.
45 Hardinges, J. and Simperl, E. (15 October 2024). "A data for AI taxonomy". Open Data Institute Blog.
46 FAIR data meets the FAIR principles of Findability, Accessibility, Interoperability and Reusability. See: https://www.go-fair.org/fair-principles/.
47 Open Data Institute. (2025). A framework for AI-ready data.
48 Jonker, A. and Aquino, J. (n.d.). "Why AI data quality is the key to AI success".
49 Ibid.
50 Fu, H. et al. (21 July 2025). "From open data to AI-ready data: Building the foundations for responsible AI in development". World Bank Blogs.
51 Ibid.
52 Dalberg Data Insights. (n.d.). The People-Centered AI Playbook.

AI for data

While data shapes AI, the reverse is also true: AI is fundamentally transforming the way data is generated, processed, managed, analysed and used. Emerging technologies have contributed to the immense volume of data being generated, for instance, through IoT devices, Generative AI (GenAI) outputs, and overall digitalisation. At the same time, AI has created new opportunities for data discoverability,[53] analysis[54] and improving the usability of unstructured, historical or analogue data.[55] Synthetic data, or data generated by AI, provides an opportunity to fill Africa's data gaps while reducing privacy violations,[56] but also comes with risks around quality, bias and "model collapse" when overused.[57] The lack of usable local datasets can lead AI developers to resort to synthetic data despite these risks.

Kenya's data ecosystem

From championing open data with the flagship Kenya Open Data Initiative (KODI),[58] to launching an Inclusive Data Charter Action Plan[59] for stronger data disaggregation and integrating citizen-generated data in the national statistics system,[60] Kenya has shown leadership in open and inclusive data practices.

Yet efforts like KODI reveal the implementation challenges of public data initiatives. The portal is now inactive due to poor data management, the lack of a legal framework, unclear institutional ownership and inadequate capacity and resources.[61,62] Global Data Barometer classified Kenya as a "medium capability" country based on its capacity to collect, manage, share and use data for the public good, revealing both the progress the country has made and the distance it still has to go.[63]

This research analyses both the barriers and opportunities in Kenya's data ecosystem, distilling recommendations for government and other stakeholders to ensure AI development is effective, impactful and equitable.

53 Solatorio, A. and Dupriez, O. (22 May 2024). "Beyond keywords: AI-driven approaches to improve data discoverability". World Bank Blogs.
54 Global Partnership for Sustainable Development Data and GIZ. (2023). Artificial Intelligence Practitioners' Guide: Kenya.
55 For instance, MezaAI is an AI-powered tool that digitalises handwritten nutrition and health records in Kenyan clinics using optical character recognition (OCR).
56 World Economic Forum. (2025). Synthetic Data: The New Data Frontier.
57 Hardinges, J. and Simperl, E. (6 February 2025). "How the data for AI ecosystem is evolving". Open Data Institute Blog.
58 Centre for Public Impact. (2016). The Kenya Open Data Initiative.
59 Ministry of Public Service, Gender, Senior Citizens Affairs and Special Programmes. (n.d.). Inclusive Data Charter Action Plan 2021–2025.
60 Bonareri Omache, S. and Kinyanjui, J. (18 January 2023). "Collaborating to integrate citizen-generated data for official reporting on Sustainable Development Goals in Kenya". Global Partnership for Sustainable Development Data.
61 Kaaniru, J. (10 April 2025). "Ethical AI Development in Kenya: The Role of Ethical Data Sourcing and Governance". Strathmore University Centre for Intellectual Property and Information Technology Law (CIPIT).
62 Open Government Partnership. (2023). Kenya Action Plan 2023–2027.
63 Global Data Barometer. (2025). 2nd Edition Report.

Focus sectors: MSMEs and public service delivery

The Kenya AI Strategy 2025–2030 prioritises eight sectors for AI applications: healthcare, education, agriculture, public service delivery, security, micro-, small and medium enterprises (MSMEs), sustainability and the creative sector.[64] This research investigates the availability of relevant data for AI use cases in two of these sectors: MSMEs and public service delivery. These sectors were chosen for their potential for impact, alignment with Kenyan government priorities and the additional value they bring to existing research.

MSMEs

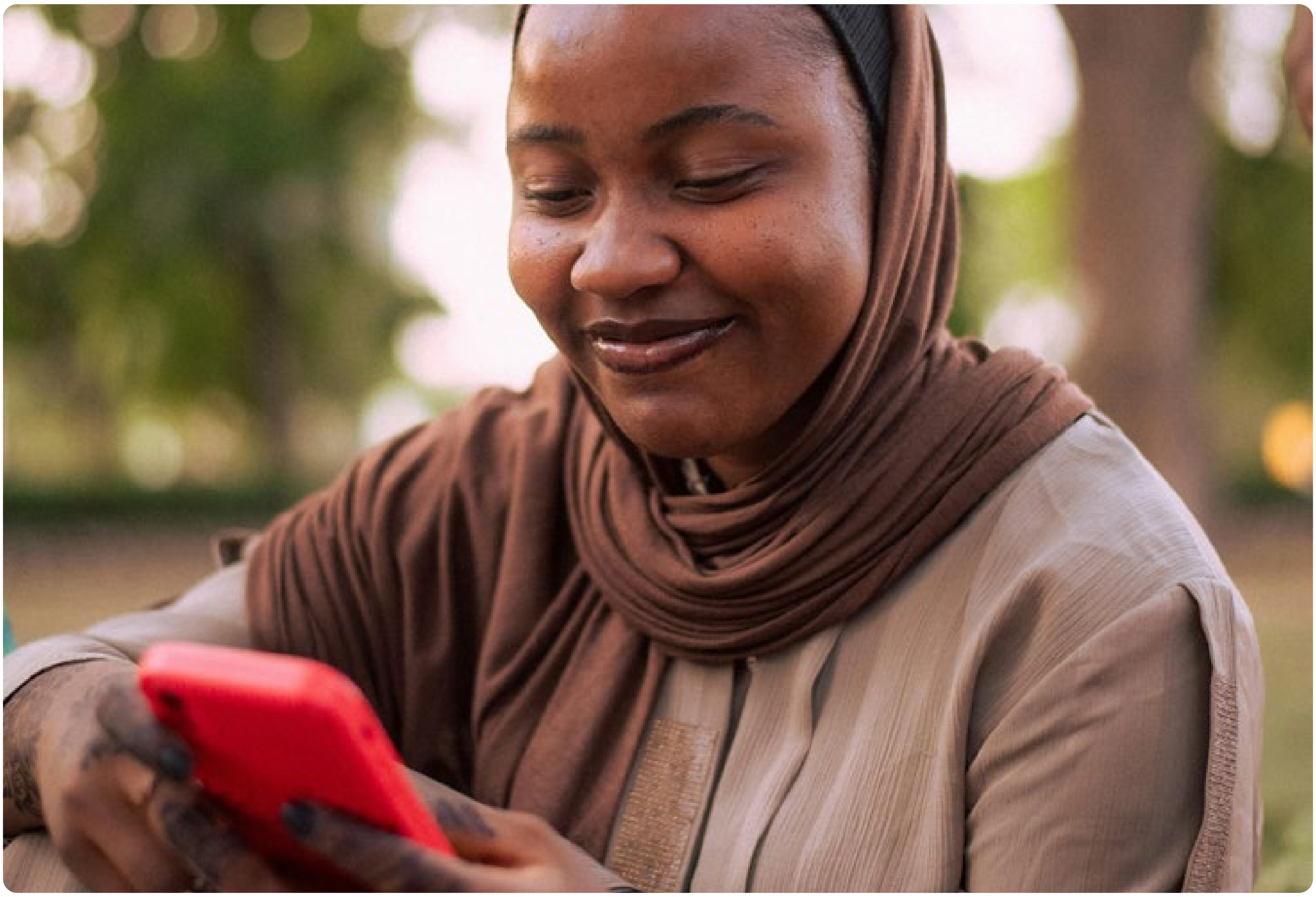
MSMEs are a key pillar of Kenya's Bottom-Up Economic Transformation Agenda (BETA),[65] contributing 40% to GDP[66] and providing about 46% of formal sector jobs and 82% of jobs in unlicensed or locally licensed businesses.[67] The national draft MSME Policy and MSME Act (2025) provide a roadmap for strengthening the sector, including the adoption of digital technologies to enhance business operations and market reach.[68] Despite Kenya's widespread connectivity and smartphone penetration, several business processes – like payments and transaction logs – remain manual, particularly for micro-enterprises that make up most licensed and unlicensed businesses. This complicates efforts to introduce advanced digital solutions to MSMEs.

40%
of Kenya's GDP contributed by MSMEs

46%
of formal sector jobs provided by MSMEs

82%
of jobs in unlicensed or locally licensed businesses supported by MSMEs

64 Ministry of ICT and the Digital Economy. (2023). Kenya Artificial Intelligence Strategy 2025–2030.
65 Government Delivery Unit. (n.d.). BETA Pillar: MSME Economy.
66 UNDP. (2021). Impact of COVID-19 on Kenyan MSMEs: Strategies for Resilience and Recovery.
67 The World Bank. (2022). Entrepreneurship Ecosystems and MSMEs in Kenya: Strengthening Businesses in the Aftermath of the Pandemic.
68 Micro and Small Enterprises Authority. (2025). "Nationwide public participation for Draft MSME Policy 2025 and Amendment Bill".



Public service delivery

Kenya has invested heavily in digitalising public services through platforms such as eCitizen, public procurement, health, education, transport, tax and identity registries, while the Digital Superhighway Agenda aims to deliver 80% of public services online.[69] The government has also laid out a Digital Public Infrastructure (DPI) roadmap, emphasising a commitment to strengthening the digital foundation of the public sector.[70] Yet public service digitalisation efforts are fragmented, contributing to a lack of interoperable and integrated data in the sector. With Kenya's devolved governance structure, its 47 county governments deliver key public services, but at varying levels of digital maturity. Advanced technological systems at the national level can mask significant variability from county to county, where the vast majority of public sector data collection occurs.

80%

of online public service delivery targeted under the Digital Superhighway Agenda

69 The National Treasury. (4 September 2025). "The Launch of the Electronic Government Procurement System".
70 Indeje, D. (15 September 2025). "Kenya Launches Strategic Roadmap to Build Inclusive Digital Public Infrastructure". KICTANet.

2. ABOUT THIS RESEARCH

Research objectives

This research aimed to meet the following objectives to strengthen Kenya's data ecosystem for AI:

Ecosystem assessment
Provide a high-level ecosystem assessment of data for AI in Kenya to ground the research.

Data assessment
Assess the status of these datasets, including perceptions of availability, accessibility, usability, shareability, localisation and quality.

Use case identification
Identify four high-impact AI use cases across two priority sectors: public service delivery and MSMEs.

Barrier identification
Identify technical and non-technical barriers in the two sectors.

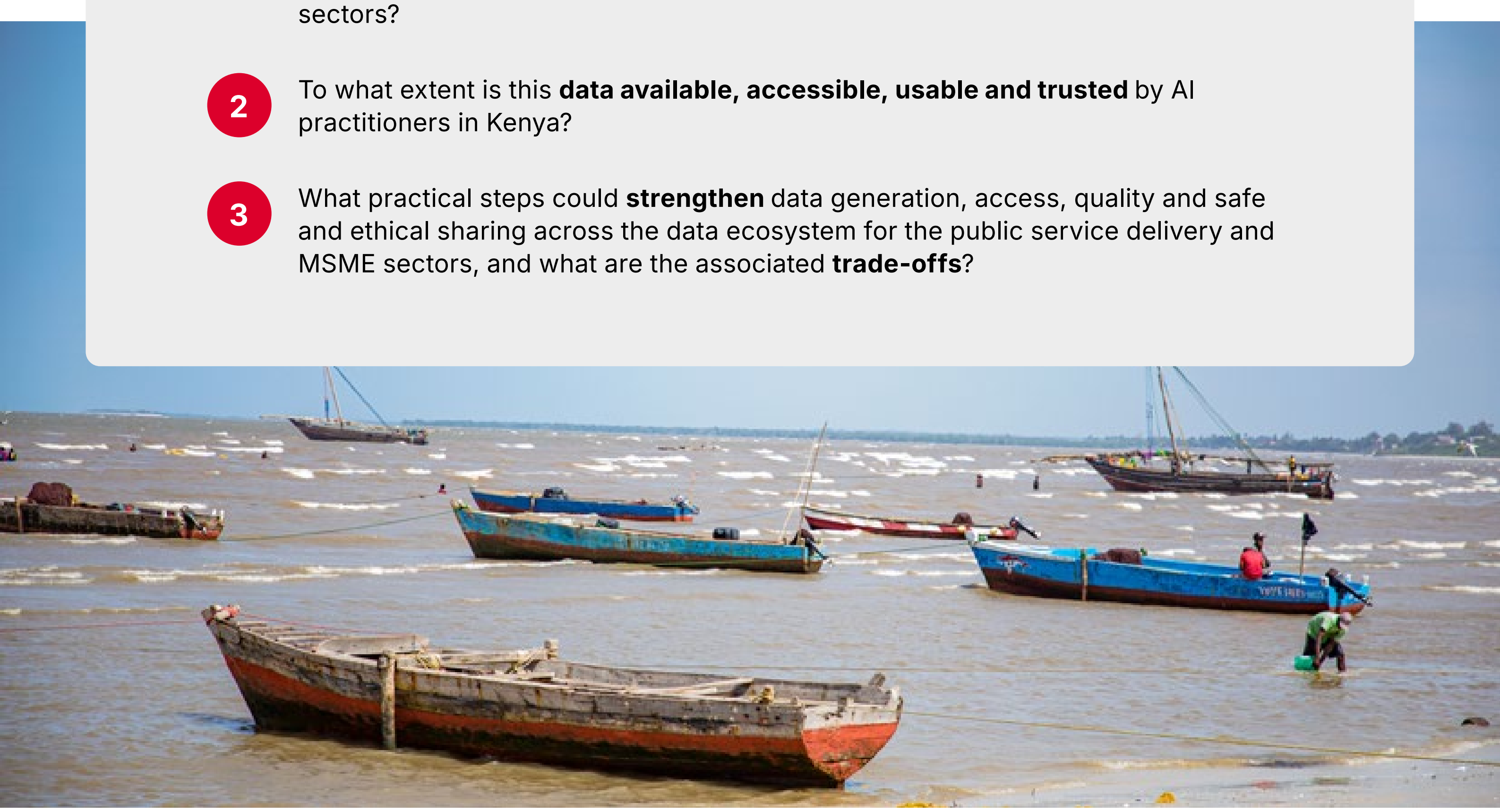
Data mapping
Map the key domain-specific and cross-cutting datasets, relevant actors and infrastructure required to support use cases in the MSME and public service delivery sectors.

Policy recommendations
Propose targeted recommendations for key actors to advance data readiness for AI.

Research questions

The study sought to answer the following key research questions:

1. What are the most relevant domain-specific and cross-cutting **data(sets) needed to enable high-value AI use cases** in the public service delivery and MSME sectors?
2. To what extent is this **data available, accessible, usable and trusted** by AI practitioners in Kenya?
3. What practical steps could **strengthen** data generation, access, quality and safe and ethical sharing across the data ecosystem for the public service delivery and MSME sectors, and what are the associated **trade-offs**?



Methodology

This study deployed a mixed methods approach, including:

Desk research and literature review

Scanned policy documents and literature in the context of Kenya, as well as two comparative countries.

Research objective met:
Ecosystem assessment, Use case identification

Key informant interviews (KIIs)

Semi-structured interviews with 43 stakeholders from the public sector (national and county), private sector, domain experts and global experts on the intersection of data and AI. Additionally, groups of experts gathered at two critical junctures: an in-person roundtable discussion midway through the project and a virtual group discussion in the final stage to test, refine and deepen the findings. Further details can be found in Appendix 1.

Research objectives met:
Use case identification, Data assessment, Barrier identification, Policy recommendations

Data inventory

Mapped data relevant to the two focus sectors, as identified by a literature scan and KIIs.

Research objectives met:
Ecosystem assessment, Data mapping

Analytical approach

Identification of use cases

Four use cases were selected to illustrate and analyse the state of data ecosystem readiness in the two focus sectors. Use cases were identified based on their **potential for impact** (the extent to which the use case benefits citizens, particularly those most vulnerable); the **added value of AI** (the extent to which AI could provide unique or transformative benefits compared to non-AI or traditional approaches); and **data feasibility** (the availability of data and existing infrastructure to support the use case, existing pilots and implementation). Note that the use cases featured in this report are neither exhaustive nor definitive, but rather serve to ground the analysis of data ecosystems in real-life examples and map the data pipelines.

Data assessment

The following analytical framework has been used to assess the datasets required for each use case. It draws from several existing frameworks: the ODI's Framework for AI-Ready Data,[71] the ISO/IEC Data Quality Model,[72] the NIST AI Risk Management Framework,[73] Data Feminism principles[74] and the IMF's AI Preparedness Index.[75] These frameworks have been consolidated and tailored to the Kenyan context and the feasibility of this research study.

Each diagnostic question within the framework was scored as: Yes/Partial/Unknown/No, with "Yes" always being a positive score. The overall assessment rating was used to guide the identification of opportunities, barriers, risks and recommendations. Because the researchers did not always have access to the raw data being studied, some details of the data (e.g. metadata) were not always possible to decipher. In these cases, KIIs and secondary research were used to arrive at an assessment.

Figure 5: Data assessment framework

1. Data generation and digitisation

Assesses whether data is routinely generated in a digital format, its coverage across national and county levels and evaluates whether the data is sufficiently granular for AI applications while ensuring diverse demographic representation.

- Is the required data being routinely generated?
- Is the data digitised in a standard format (vs paper-based)?
- Is the coverage wide (national, county, facility-level) and granular (citizen, facility, transaction-level) enough for the use case?
- Does it represent underrepresented groups? (e.g. women, youth, low-connectivity groups)?
- Are data labels routinely generated if needed?

2. Data quality and reliability

Measures the reliability of the data by checking for completeness, accuracy and consistency across different time periods, and identifies any significant biases or errors that would affect the validity of AI outcomes.

- Is the data/data fields complete (no missing values, blank entries)?
- Is the data accurate (considering verification processes, error rates)?
- Is the data timely (update lag, real-time vs batch)?
- Is the data consistent over time?

71 See: ODI AI-Ready Data Framework.
72 ISO. (2008). ISO/IEC 25012:2008 Software engineering – Software product Quality Requirements and Evaluation (SQuaRE) – Data quality model.
73 NIST. (n.d.). AI Risk Management Framework.
74 D'Ignazio, C. and Klein, L. (n.d.). "The Seven Principles of Data Feminism". Responsible Data.
75 International Monetary Fund (IMF). (2023). AI Preparedness Index (AIPI).

3. Data integration and interoperability

Examines the unique identifiers for data and the ability of different systems to link and share information, following set standards.

- Are there unique identifiers for each record (e.g. record ID)?
- Do structures that allow for shareability exist (e.g. APIs)?
- Does the data follow standardisation that allows for interoperability?

4. Data governance and access

Evaluates compliance with existing legislation, the presence of formal data-sharing agreements between institutions and assesses whether clear ethical safeguards and access controls are defined to ensure responsible stewardship.

- Is there compliance with the Kenya Data Protection Act and Access to Information Act?
- Are data-sharing agreements or MoUs in place?
- Are ethical safeguards defined (fairness, explainability)?
- Is there clarity on who can access the data?

5. Data infrastructure and compute

Assesses the technical capacity for storage and processing, and the availability of necessary analytical tools at the institutional level.

- Are there data storage systems (cloud/hybrid) to support the use case?
- Do data systems store both structured and unstructured data (e.g. text, images, video)?
- Is there compute capacity for machine learning (ML) (e.g. GPUs/TPUs, cluster access)?
- Is there connectivity at various levels (e.g. national and county level)?
- Are data modelling tools available (e.g. ETL, analytics, ML platforms)?

6. User and use case alignment and sector readiness

Verifies if the currently available datasets match the specific technical requirements needed to train the proposed AI models and measures the readiness of key actors to adopt solutions in a real-world setting.

- Do current datasets match the minimum data needed for the AI use case?
- Are there gaps that block model training (e.g. not enough images, no historical logs)?
- Are frontline staff or county teams prepared to use or generate the required data?
- Are there existing pilots or proof-of-concepts?
- Do key actors have the skills and/or trust to engage with the use case?

3. MAPPING KENYA'S DATA ECOSYSTEM

This section provides a detailed overview of Kenya's data ecosystem and the barriers, drawing on KIIs and secondary research.

Data and datasets

Datasets and data assets refer to the data resources that are available or will be available to support the training, testing, development and deployment of AI systems. These systems use a mix of data, derived from a wide range of sources: from websites to publicly available research, personal data to Internet of Things (IoT) sensors. Figure 6 describes some of the key data categories relevant to AI development in Kenya, which vary considerably by accessibility, quality and readiness for AI applications. In addition, domain- and sector-specific data assets relevant to AI applications in the MSME and public service sectors will be mapped in detail in the next sections.

Figure 6: Data relevant to AI development in Kenya

Statistical data

National statistical data gathered through censuses and surveys covering population demographics, economic indicators and sector-specific statistics

KEY ACTORS/ INITIATIVES

Kenya National Bureau of Statistics (KNBS)

County governments

AI APPLICATIONS SUPPORTED

Policy modelling, economic forecasting and public service optimisation

BARRIERS

- Limited resources and respondent fatigue
- Lack of timeliness and regularity (long gaps and delays between survey/census editions)

Administrative data

Data generated through government operations (e.g. human resource records, monitoring and evaluation (M&E) reports, licences and registries)

KEY ACTORS/ INITIATIVES

County governments

eCitizen

Kenya Revenue Authority (KRA)

Ministry of Land and Physical Planning (MoLPP)

National Land Commission

AI APPLICATIONS SUPPORTED

Policy modelling, economic forecasting and public service optimisation, revenue management

BARRIERS

- Fragmentation
- Limited digitalisation
- Not in AI-ready formats
- Poor sharing mechanisms

Government · Research & academia · Civil society / NGOs · Private sector · International / multilateral

Language data

Curated text or speech datasets representing the linguistic and communication patterns of a language

KEY ACTORS/ INITIATIVES

KenCorpus[76] | African Next Voices Initiative[77]

Masakhane[78] | Mozilla Common Voice[79] | Lacuna Fund[80]

AI APPLICATIONS SUPPORTED

Natural language processing (NLP) applications including speech recognition, translation systems, conversational agents, AI tools that can operate effectively in local contexts, citizen-facing use cases (e.g. multilingual chatbots)

BARRIERS

- Inadequate representation of languages beyond Swahili and transboundary languages, leaving more than 40 Kenyan languages "low resource"[81]
- Shortage of data labelling and annotating skills

Geospatial data

Time-based data that relates to a specific location on or near the earth's surface, collected through satellites, sensors, surveyors, GIS and GPS systems[82]

KEY ACTORS/ INITIATIVES

Kenya Meteorological Department

Directorate of Resource Surveys and Remote Sensing

Ministry of Land | Kenya Forest Service and Kenya Wildlife

Kenya Space Agency | Council of Governors' GIS dashboards[83]

Kenya Agricultural and Livestock Research Organisation (KALRO)[84]

Kenya Agricultural Observatory Platform (KAOP)

Digital Earth Africa | Kenya Flying Labs | NASA Harvest

UNOCHA Humanitarian Data Exchange (HDX) platform[85]

Microsoft AI for Good Lab | OpenStreetMap (OSM) Communities[86]

AI APPLICATIONS SUPPORTED

Geospatial analytics in sectors such as agriculture, climate monitoring, spatial and urban planning, disaster response and public service delivery

BARRIERS

- Fragmentation and duplication
- Cost of access

Government | Research & academia | Civil society / NGOs | Private sector | International / multilateral

76 See: KenCorpus
77 See: African Next Voices
78 See: Masakhane
79 See: Mozilla Common Voice
80 See: "waleghwa/low-resource-language-data: Parallel Corpora for Kiswahili and Kidaw'ida, Kalenjin and Dholuo"
81 Amol, C.J. et al. (2024). "State of NLP in Kenya: A survey". arXiv.
82 IBM. (n.d.). "What is geospatial data?"
83 Maarifa Centre. (n.d.). "GIS Dashboards & Data".
84 Appiah Acquaye, N. (13 February 2026). "Kenya launches CRoME Initiative to boost agriculture with space technology and AI". Tech Review Africa.
85 See: The Humanitarian Data Exchange
86 See: OpenStreetMap

Barriers

Challenges with data digitisation and standardisation compromise the quality and accessibility of data in the public service delivery and MSME sectors, as well as its usability for AI solutions.

0101 1001 Digitisation

- Despite progress in data digitisation in the public sector, **the continued use of manual and paper-based data processes** in national and county governments is a barrier to data access, sharing and use in AI systems. Digitisation is a time- and resource-intensive process that **deters local innovators and researchers from even attempting to access county-level data.**

> "We went through a long process of signing NDAs etc [with a national ministry], and they ended up giving the data in flash discs and hard drives, which beats the logic at the end of the day."
> **– Innovator**

- Even when data collection is digitised, it often **replicates existing physical processes rather than automating them** (for instance, Excel or paper-based templates that are later uploaded to computers), **leading to incomplete and out-of-date datasets.**
- When government statistical or administrative data is made available, it is often in **formats that are not optimised for machine-readability**, such as PDF reports. Government departments like KNBS are focusing on making statistical data open and "AI-ready" to be used in models, but most other departments lag behind. Also, while KNBS makes open data available in digital formats, these datasets are often aggregated to maintain confidentiality, lacking the granularity needed for some AI applications.

Standardisation

- Across the data value chain, **a lack of standardisation in collection approaches, formats, standards, metrics and definitions precludes the development of advanced AI solutions that integrate different kinds of data.** For example, the Kenya Statistical Quality Assessment Framework (KeSQAF)[87] provides standardised guidelines for data collection, management, processing and dissemination, but there is a lack of consistent, nationwide adoption of these standards by government departments and other sectors due to low awareness, difficulty putting frameworks into practice and limited capacity in advanced data science.

> "Even something like the format of collecting dates is not standardised, so you can't use the data meaningfully as a result."
> **– Data scientist**

- **Without clear quality assurance mechanisms and evaluation processes, assessments of data quality are reactive rather than proactive.** Without dedicated efforts to identify gaps in datasets, quality issues persist in new data collection efforts.

> "We really don't have metrics for measuring the quality of data. Whenever a challenge arises that is when we go back and try and evaluate how accurate that data was."
> **– County government official**

- Poor standardisation necessitates **extensive data cleaning** when creating regional or national AI models, which adds costs for smaller innovators.
- Researchers or AI developers can be **hesitant to use data collected and shared by other actors**, including the private sector, due to a lack of transparency on data processes and an inability to ascertain the quality of the data.

87 KNBS. (2023). Kenya Statistical Quality Assurance Framework (KSQAF).

Data collection approaches

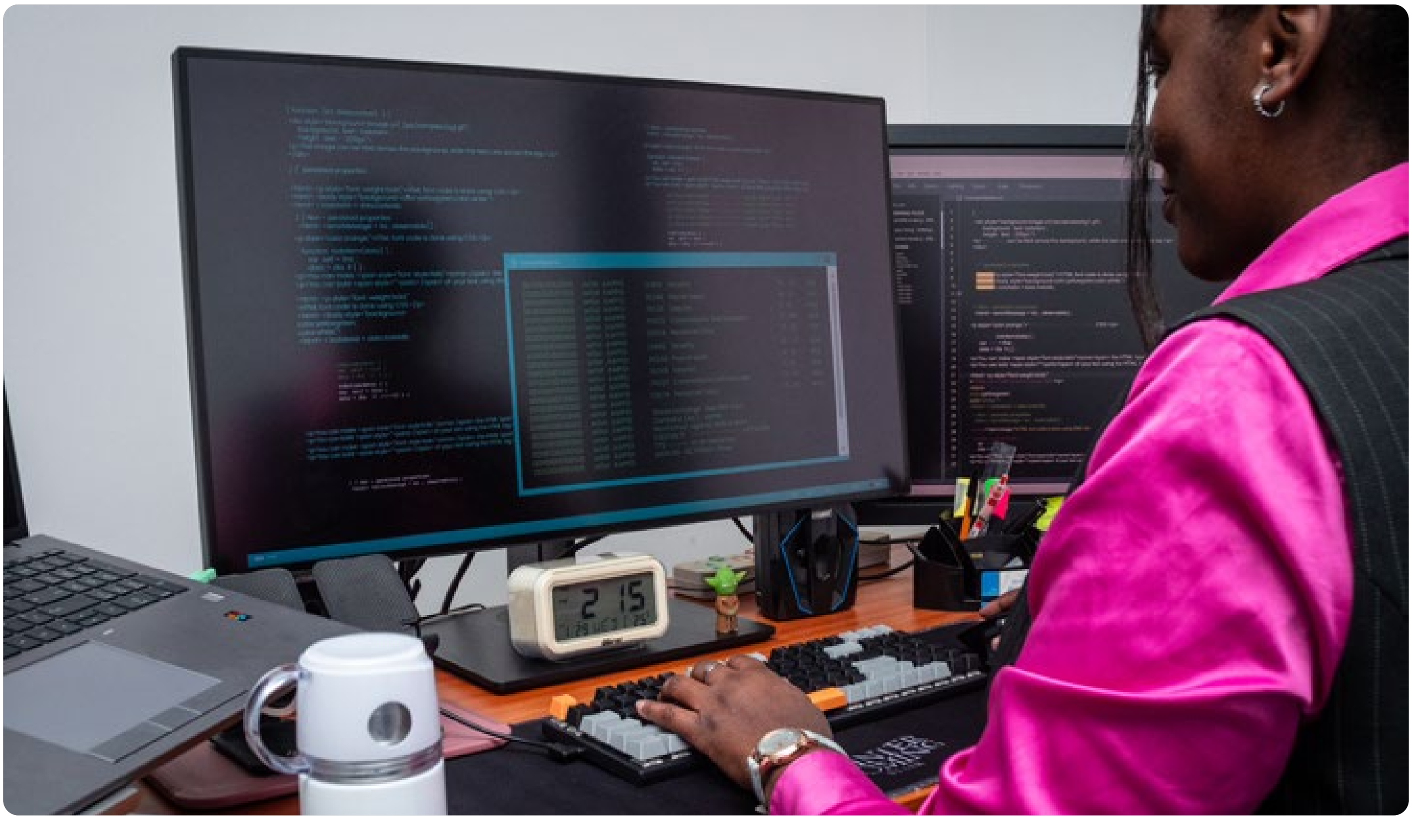
Given these barriers to accessing clean, high-quality, digital data, local startups, innovators and organisations in the public service delivery and MSME sectors are **building their own data pipelines appropriate to their context, leveraging their strong understanding of users and established trust.** Smaller organisations often prefer community- and citizen-based approaches to data collection, and have deployed accessible channels like SMS, USSD and IVR, apps with offline capabilities, WhatsApp groups with farmers and community champions.

In addition to being more collaborative, crowdsourced data can also be more sustainable than, for instance, expensive IoT or satellite data collection. However, crowdsourcing has its own **financial and operational costs.** For instance, when the Local Development Research Institute (LDRI) worked with farmers to gather data for an early warning system, they had to provide smartphones with high-quality cameras and geolocation features to ensure farmers collected accurate and usable data. Crowdsourced data also requires **strong validation and quality-control layers,** such as constantly monitoring and triangulating data, defining clear internal thresholds, conducting spot checks and training data collectors.

At the same time, **innovators are increasingly deploying emerging technologies to collect, digitise and make sense of data.** Startups are using AI to clean up unstructured SMS survey responses, optical character recognition (OCR)[88] to digitise and produce insights from PDF documents or scans, apps that passively collect data from a mobile device with users' consent, social listening on social media to gather citizen perspectives and synthetic data for health use cases. This points to a strong use case for AI and other emerging technologies: making data available, clean and usable before it is deployed. While these digital tools could add significant value in the public sector, they are often out of reach due to a lack of resources, capabilities or awareness.

While startups invest in primary data collection because of challenges accessing high-quality data, **they are often reluctant to share their datasets.** Building tailored data pipelines provides innovators with a strong competitive advantage, but data "moats" can also slow ecosystem-wide collaboration. It is rarely data alone that differentiates a startup, but rather the combination of data, the AI model and its ultimate application.

88 OCR is a technology that extracts data from images, scanned documents or PDF texts and turns it into a machine-readable format. OCR can integrate AI for more advanced character recognition. Source: IBM.



Data management and storage systems

The digitalisation of government services in Kenya's public sector has begun to transform the way administrative data is collected and stored. With the devolution of sectors like agriculture and health, counties are important, albeit fragmented, hubs of localised data, with data systems at varying levels of digital maturity.

eCitizen is a centralised administrative dataset with potential to support improved service delivery, data analytics and AI-driven innovations in the public sector. The Government of Kenya has also deployed several management information systems (MIS) to enhance efficiency, transparency and accountability in administrative functions across sectors, such as the e-Government Procurement System (eGP), Integrated Financial Management Information System (IFMIS), Kenya Agricultural Information Management System (KIAMIS), Health Management Information Systems (HMIS), revenue and tax management systems and ministry-specific platforms like the Ardhisasa platform,[89] developed by the Ministry of Land and Physical Planning (MoLPP) and the National Land Commission (NLC) to digitalise land records.

However, fragmented, legacy systems remain siloed in government departments or agencies, complicating efforts to link datasets or establish automated data-sharing mechanisms between public sector institutions. This is also the case at the county level, where even digitally mature counties experience data duplication and inconsistencies across departments. For instance, Murang'a County uses a cloud-based system for its revenue data, but in the absence of a unified master system, linking different datasets still requires manual matching of IDs or locations via spreadsheets and GIS overlays (where multiple spatial data layers, such as land use, roads and topography sharing the same coordinate system, are combined) to analyse interrelationships.

89 See: Ardhisasa

Despite advancements in public sector digital systems, **many were designed for specific institutional functions that are no longer compatible with modern data analytics or AI applications, and were not designed with interoperability standards in mind.** Upgrading or modifying existing data systems to comply with new technological standards, for instance, integrating them in interoperable digital ecosystems or centralised API gateways, comes with high retrofitting costs.

Vendor lock-in is another recurring challenge for data systems in national and county governments. Vendor contracts may lack explicit definitions of who owns the application and the underlying database once contracts end, leading to high retrofitting costs, data privacy vulnerabilities and the risk of departments losing access to their own data. Legacy systems built by competing vendors can even prevent data sharing between agencies within the same ministry.

Without centralised, searchable databases, data sharing in the public sector is sometimes a manual process, through emails or meetings. For instance, in one county, a weekly Director's Caucus meeting serves as the mechanism to validate and share information and progress across the county's various departments. KNBS uses spreadsheet templates to gather monthly data from all county departments to compile county statistical abstracts. This not only slows data sharing but can compromise the security and longevity of data.

Promising efforts are being made to address interoperability challenges in data management and governance and make datasets more accessible and discoverable. The Ministry of ICT's API Gateway Strategy aims to facilitate efficient and secure data exchange across government entities. Its GAVA Connect platform is currently being piloted as a centralised API gateway that can streamline and simplify data access and sharing and support the integration of AI technologies within government systems. The ICT Authority's Government Enterprise Architecture (GEA)[90] framework aims to standardise ICT systems, data management and interoperability across public institutions, while the ICT Authority Strategic Plan (2024–2027) aims to revitalise the KODI platform to improve data access policies, enhance transparency and support the development of data-driven applications, including AI solutions. New DPI like the Agriculture Sector Data Gateway (ASDG)[91] are being developed to centralise fragmented agricultural data in an interoperable data exchange. **Universities, non-profits and multilateral organisations** are also working to improve the accessibility and discoverability of African datasets, for instance, with the AfriData platform,[92] the Integrated Data Atlas[93] and the National Economic Map.[94] **County governments are also pursuing digitalisation and open data initiatives,** such as Nandi and Kilifi Counties'[95] open data portals. These efforts need to continue to be supported at an ecosystem level.

90 ICT Authority (ICTA). (n.d.). "About GEA ICT Standards".
91 See: Ministry of Agriculture & Livestock Development
92 Developed by JHUB at Jomo Kenyatta University of Agriculture and Technology (JKUAT).
93 A data platform developed by the Institute for Development Data to access reference data, live signals, historical indicators, markets, trade and government records for all African countries through a single API. See: https://atlas.data4dev.org/about.
94 Digitalises, analyses and visualises public data across all Kenyan counties: https://opendata.co.ke/
95 See: Kilifi Open County Data

While access to government data might not be perfect, there are data sharing and access pathways that could be strengthened. The following examples highlight the conditions that can make effective data sharing possible.

Clearly defined mandates

To complement its own data collection efforts and fill information gaps, KNBS accesses data from other government departments, for instance, data on crop production from the State Department for Agriculture and data on imports from the Kenya Revenue Authority (KRA), and combines it with household consumption data.

Institutionalised processes and structures for anonymisation

KNBS shares raw, anonymised datasets with researchers through formal request procedures authorised by the Director General, rigorously anonymising data to ensure individuals cannot be identified. Ushahidi also has a framework for sharing its crowdsourced citizen data with noncommercial researchers through a standardised data-sharing agreement on its website, ensuring all personal identifiable information is redacted at source.

Crisis and humanitarian response settings

During the El Niño floods, Action Lab brought together 18 different organisations, including private entities like Uber, Meta, Safaricom and Planet Data, to aggregate data and support the Red Cross in proactive decision-making.

Government support and buy-in

As part of a universal health coverage pilot led by the Kenyan government and the World Bank, an insurtech (insurance tech) shared data to support the mapping of counties. While a lack of governance frameworks has been a barrier for private players in other cases, the strong backing and push from the government was a key enabler in this instance.

Long-term embedded partnerships and trust

In the public sector, innovators have strengthened data sharing within government departments by acting as embedded partners driving systems-level change, rather than as commercial vendors. For instance, Action Lab spent several months collaborating with the judiciary to fundamentally rethink its data infrastructure and build trust before introducing AI models. Establishing an underlying digital data management pipeline allowed for more advanced data practices to be built on top. In another instance, Machakos County built and deployed an AI supervisory tool for community health workers in partnership with Dalberg Data Insights. Over nearly a year, the group worked to build trust, engage government stakeholders and establish local ownership before technical deployment, ensuring the county felt secure in sharing and managing data, and that they were plugging into existing electronic community health systems.

Leveraging data protection laws

As part of Pathways Technologies' work developing digital solutions for the public sector, stakeholders often cite concerns with data sharing, impeding progress. Pathways applies a close reading of the DPA to educate stakeholders about legal provisions and demonstrate their compliance, such as through their use of sovereign government data centres (like KAIST) and ensuring a copy of the data remains on-site, enabling ethical data sharing.

Data actors

Data and data ecosystems are fundamentally shaped by the individuals, communities, organisations and institutions that create, contribute to and benefit from data. Data value chains depend on collaboration and aligned incentives and data itself ultimately represents people and communities. Human behaviour, motivations, relationships, capabilities and institutional dynamics therefore all play a major role in ensuring data is collected, managed and used effectively and collaboratively.

Table 1: Roles in the data ecosystem

Data for AI ecosystem role	Examples
Contribute to data (as data subjects)	Individuals and communities, including farmers, micro entrepreneurs, service users
Hold data (institutions that generate or collect primary data from users, devices, or environments)	Financial institutions, MNOs, Big Tech companies, government agencies, health facilities
Process data (including cleaning, labelling, annotating data to make it AI-usable and moderating or assessing data during model training and evaluation)	Startups, innovators, research institutions, data workers
Fund data (including data insfrastructure, partnerships, tools and systems across the value chain)	Governments, development partners, Big Tech companies
Deploy data (institutions that put data to use for AI solutions)	Model developers like Big Tech companies, frontier labs, research institutes; application developers like startups, innovators, research institutions, domain specialists
Impacted by data (in addition to those who directly deploy data, the individuals and institutions who are impacted as end users or indirect users of AI-enabled systems through B2B, B2C, B2G deployments)	Citizens impacted by AI-enabled public services or policy decisions, farmers using an AI advisory service, small or medium enterprises, financial institutions, MNOs, government agencies

Role of the private sector

The private sector in Kenya plays a central role as data generators and holders, while also driving broader AI development in the country. This is often done in collaboration with the public and civil sector, indicating the importance of multistakeholder collaboration in strengthening the AI data ecosystem.

The private sector, including financial institutions, Big Tech, e-commerce companies and MNOs, holds vast amounts of digitalised, granular data that can be valuable for robust AI training and applications, particularly in the MSME sector. However, these proprietary datasets are often inaccessible or too expensive for civil society, academia and local innovators, and data sharing is complicated by commercial sensitivities and regulations.[96,97] This gives players like global tech companies a competitive advantage when developing AI solutions, while raising criticisms of the disproportionate benefits companies extract from African data and labour.[98]

In addition, the private sector drives much of the country's AI development[99] through research and innovation relevant to public service delivery (e.g. Google AI lab, Microsoft Africa Development Centre in Nairobi), capacity building for the public sector and MSMEs (e.g. Microsoft and Kenya Private Sector Alliance (KEPSA)'s Kenya Artificial Intelligence Skilling Alliance (KAISA)) and infrastructure development (e.g. IXAfrica data centre, the Olkaria EcoCloud hub with investment from G42[100] and Google's Umoja subsea fibre optic cable[101]).

Partnerships are at the heart of many of these successful initiatives. The Crop Measurement and Evaluation dataset initiative, a collaboration between KALRO and Microsoft,[102] shows the value of partnerships in mobilising resources, improving data collection and expanding access to datasets. However, some public stakeholders might be hesitant to become dependent on private datasets, as there is a risk of unexpected termination, costs being raised or low interoperability. Identifying mutually beneficial use cases, building long-term sustainability plans and collaboratively developing a data framework to manage every step of the data value chain are important factors to consider.[103]

96 Orembo, L. (6 September 2025). "Data Deficits and Democratic Processes in Africa". African Alliance for Access to Data.
97 Ibid.
98 Chetty, P. and Sey, A. (10 October 20256). "RIA Just AI Framework of Inquiry". Research ICT Africa.
99 Google and Public First. (n.d.). The Digital Opportunity of Kenya.
100 G42. (6 March 2024). "G42 and Kenya's Ecocloud unveil green-powered mega data center collaboration".
101 Quigley, B. (23 May 2024). "Improving connectivity and accelerating economic growth across Africa with new investments". Google Cloud Blog.
102 Appiah Acquaye, N. (13 February 2026). "Kenya launches CRoME Initiative to boost agriculture with space technology and AI". Tech Review Africa.
103 GSMA. (2021). Innovative Data for Urban Planning: The opportunities and challenges associated with public-private data partnerhips.

SPOTLIGHT:
The role of MNOs in the data ecosystem

Data and datasets

Given Kenya's high rate of mobile penetration, MNOs hold vast quantities of data, including:[104]

- **Event data:** Logs recorded by an MNO when users connect to their network for calls, SMS, mobile internet, USSD, mobile money transactions or other type of event data recorded by the MNO's system.
  - Mobile money transaction data is a standout in Kenya. Safaricom's M-Pesa serves as the backbone of Kenya's digital economy, with 40 million customers, 2.4 billion businesses accepting M-Pesa payments and an entire ecosystem of transfer, payment, credit, loan and business offerings.
- **Network data:** Data on the telecoms network, including the location of cells, antennas and underground networks, infrastructure status and other logs of activities taking place on the network.
- **Customer data:** Includes socio-economic and demographic information, as well as information related to customer sign-up, status and activity.

Infrastructure

After helping to lay the foundation of Kenya's information and communications technology (ICT) infrastructure, MNOs are now contributing to national AI infrastructure through data centres[105,106] and undersea cables.[107]

Data systems

Safaricom has developed several public sector digital data platforms: the E-fertilizer Subsidy Management System developed for the Ministry of Agriculture to disburse fertiliser vouchers;[108] myCounty app, developed in collaboration with Makueni County to streamline revenue collection and management at the county level; and health information systems.[109]

MNO data sharing

Regulatory factors influence data-sharing partnerships. Following strong data privacy guidelines, establishing limitations on sharing and ensuring data is only used for its intended impact can enable ethical and secure data sharing.[110] There are also commercial trade-offs that prevent MNOs from making their data available. The GSMA has identified different business models (Figure 7) that can make data sharing with the public sector more sustainable and scalable.[111] GSMA research has also identified additional enabling factors for data-driven public-private partnerships (PPPs).[112]

Partnerships between startups and MNOs have been successful and impactful when there are structured and diverse ways to collaborate, such as sandboxes, pre-approved API frameworks or fast-track compliance pathways, and they are grounded in co-creation, co-development and shared goals for long-term sustainability.[113]

Figure 7: A typology of mobile big data partnership business models

Philanthropy	Donor-subsidised	Cross-subsidised	Tiered	Commercial
Ad hoc or continuous projects with pro bono data	Costs covered by external donor funding	Ad hoc or continuous revenue from other big data projects	Free/trial-based with some elements paid for	Full paid-for product

Source: GSMA, 2018

104 Ibid.
105 Tanner, J. (14 May 2025). "Safaricom partners with iXAfrica for hyperscale data centre solutions". Developing Telecoms.
106 Ma, J. (12 September 2025). "India's Airtel breaks ground on 44MW data center in Kenya". Data Center Dynamics.
107 Faikya, V. (28 January 2025). "Safaricom seeks approval for Kenya's first telco-owned submarine cable amid rising competition". Techpoint Africa.
108 Safaricom. (9 November 2021). "Ministry of Agriculture Selects Safaricom to Streamline Subsidy Voucher Program". Press release.
109 Safaricom. (2025). "Enterprise and Public Sector". Safaricom PLC 2025 Annual Report and Financial Statements.
110 GSMA. (2021). Innovative Data for Urban Planning: The opportunities and challenges associated with public-private data partnerhips.
111 GSMA. (6 February 2019). "Big Data for Social Good Toolkit Overview".
112 GSMA. (2021). Innovative Data for Urban Planning: The opportunities and challenges associated with public-private data partnerhips.
113 GSMA. (2026). Stronger Together: Partnerships between mobile network operators and startups for impact at scale.

Individual and institutional barriers

In the absence of institutionalised data processes, individual relationships and political priorities play an outsized role in the data value chain. Ad hoc data practices that rely on individual priorities can have both positive and negative impacts on the data ecosystem, but are ultimately volatile and regularly disrupt established partnerships, systems and practices.

Funding practices:

- **Donor-funded systems cause interoperability and data-sharing challenges.** Several data systems and tools in Kenya's public sector are donor funded, including software for data collection and management, access to satellite data and one-off projects that involve data gathering. This donor support is valuable in the short term given the limited resources of government institutions and civil society organisations to invest in data systems. **However, newly introduced donor-funded systems or platforms often do not integrate with existing systems,** limiting data sharing across counties or even within a single organisation.
- **Donor-funded systems can also be difficult to sustain once project funding ends, disrupting data pipelines and impeding progress.** For instance, a geospatial data collection platform funded by an international development partner for a national government agency was no longer usable once funding concluded. A pilot for the Agriculture Sector Data Gateway (ASDG), designed to centralise national agricultural data, was halted when donor funding was lost, leaving the fully developed platform unimplemented.

Incentives:

- **Data practices in government institutions are geared towards existing political priorities,** with data collection seen as an end in itself rather than a means to an end. Counties, for instance, focus on collecting data for their reporting, revenue-setting and compliance objectives, using data systems designed for these objectives. They have little incentive to consider how that data could be useful for broader data-driven decision-making or AI solutions.
- **For actors to improve data practices and systems, they need to be convinced of the benefits. Yet the benefits of data are often invisible to actors across the data value chain or do not capture their attention in the same way as other, more visible layers of the AI ecosystem.** For instance, farmers and frontline data collectors might not see how their farm-level data collection efforts impact decisions like agriculture resource management.

  However, because AI applications in the public service delivery and MSME sectors are still nascent in Kenya, there is a lack of high-quality evidence, making it **difficult to build a shared understanding of the benefits of a strong data ecosystem** and the willingness to invest in it. Success has come from demonstrating the real-world value of data and making incremental progress with the data available, tying data practices to how it will ultimately be used and, at the county government level, aligning incentives with their existing priorities.
- **Individual actors in government agencies can control or limit data sharing** based on political motivations or concerns about budget ramifications and scrutiny, while on the flipside, **strong data champions within county governments have been able to drive digitalisation and data improvement efforts.** Data processes and systems often last only as long as a government's tenure, and both progress and historical data get lost in the process.

Relationships and collaboration:

- **Stakeholder and partnership management,** more than any technical process, is one of the biggest challenges faced by those developing AI solutions. Accessing data from governments **hinges on personal relationships and ambiguous processes** – MoUs, access letters and templates – with varying chains of command that differ from one department, agency and instance to another.
- The **pace** at which these processes move is much slower than AI innovators or researchers would like, which can discourage innovation in the public sector in particular.
- **County governments play an important role by brokering connections between data collectors and community stakeholders.** Even when organisations or individuals are collecting primary data themselves for an AI use case, they need authorisation and permits from county governments to do so, highlighting their significant role and the importance of collaboration.

> "When we started working with [the county], it had a different government, and by the time we were building a model, the new government did not want to continue. It was really difficult building new relationships and showing the value that we were there to provide."
> **– Innovator**

Capacity, skills and trust

National and county governments are increasingly recognising the need to improve their in-house data skills. This is particularly important as AI increases the complexity of the data life cycle.[114] For instance, KNBS has a small but growing data science team that deploys ML and advanced data science to harness alternative datasets and improve the quantity and quality of their own data. Murang'a County mandates staff training on data accuracy, anonymisation and protection guidelines as a prerequisite for AI implementation, while also building internal capacity through internships for university students. Kilifi County has partnered with Power Learn Project (a nonprofit building young people's digital skills) to employ trained students in county digitalisation efforts. **However, making county governments attractive professional destinations for young people with strong digital skills remains a barrier.**

> "We were less than eight ICT staff in the county before implementing our AI-GIS system. We realised we needed to build our capacity."
> **– County government official**

While AI and data science-related skills are newer capabilities, basic data collection skills need to be strengthened. For instance, in AT4D and UNICEF's work, a lack of trained researchers among frontline workers collecting data on children with disabilities led to biased data, while stigma and misunderstanding of disability also contributed to underreporting. This highlights the potential consequences of a lack of capacity, and the importance of strong data skills when working with public service delivery data in particular, to ensure vulnerable populations are fairly and accurately represented. Communities and frontline workers increasingly play a role in data collection, too. Nonprofit Humanitarian OpenStreetMap Team (HOT) supports volunteers to collect GIS data, without the need for technical skills. Despite simplifying the process, collecting geospatial data remains complex, and while the organisation provides virtual and in-person capacity building, this can be operationally and financially taxing.

Alongside digital skills and capacity, building the trust of data subjects is key to safe and responsible data collection. In the public service delivery sector, for instance, individuals or organisations working with farmers must assure them that the hyperlocal data they are collecting will not be shared with or misused by government authorities. Similarly, informal businesses worry that their data can be shared with tax authorities, reducing their willingness to be part of data collection efforts. While Kenya's Data Protection Act (DPA) protects the privacy and safety of citizens' personal data, understanding of the Act is low in rural counties.[115]

Ushahidi, a nonprofit building open-source citizen data crowdsourcing tools, stands out in their trust-building approach to data collection. It factors in data protection right at the source, leaving out details like full name or ID number if they do not add value to how the information will be processed or used. Before beginning a data collection exercise in a community, they work with a community leader and conduct onboarding activities to help people understand the purpose of the data collection, ask questions and discuss their concerns and provide a point of contact in case of any privacy violations. This trust also paves the way for trust in the AI solutions that are built on the data collected.

Trust is just as important for those at the other end of data collection – frontline workers, field officers or community health workers – especially when introducing new digital data practices and systems. Providing sufficient training and ensuring troubleshooting support is available helps build the trust of those using digital data collection tools.

114 Global Partnership for Sustainable Development Data and GIZ. (2023). Artificial Intelligence Practitioners' Guide: Kenya.
115 Amnesty International. (October 2025). 5 Years On: Citizens' Perspectives on Kenya's Data Protection Act Implementation.

SPOTLIGHT:
Kenya's data workers

Data workers perform essential but overlooked tasks within the data and AI ecosystem, such as data labelling, annotation, generation, refining and algorithmic training. Kenya, like other African countries, is a major source of labour for data companies driven by a young, digitally savvy, English-literate population facing high unemployment rates. Kenya has also seen legal challenges to violations of data workers' rights, such as low pay, harmful and precarious working conditions and exposure to highly graphic, traumatising content with little recourse or safeguards.[116]

As AI advances automate basic annotation tasks, the qualifications required for data work are becoming more demanding, yet there is a ceiling on workers' opportunities and growth. Data workers represent the invisible human side of data processes, AI systems and the unethical data practices fuelled by rapid Big Tech AI development. Consulting data workers on data policies and amplifying their demands, such as the Kenya-based Data Labelers Association (DLA) code of conduct,[117] are essential for a just and fair data ecosystem.

116 Cockerell, I. (22 October 2024). "Legendary Kenyan lawyer takes on Meta and Chat GPT". .coda.
117 See: Data Labelers Association

Data governance

Data governance involves the technical, policy and regulatory frameworks used to manage data across its life cycle.[118] In Kenya, compliance with and enforcement of data regulation, regulatory grey areas, and a lack of clear understanding and consistent interpretation are all governance challenges in the data ecosystem.

While Kenya is developing new legal frameworks specific to AI, like the Artificial Intelligence Bill (2026) and the KEBS Draft Code of Practice for AI Applications,[119] the country already has a strong technology policy and legal ecosystem to build on when governing data and AI, as detailed in Figure 8.[120] The most significant data-related legislation is currently the 2019 Data Protection Act (DPA), which defines clear rules on data sovereignty and the ethical management of personal information. Section 35 of the Act provides citizens the right to object to decisions based solely on automated processing, which could be interpreted as mandating human-in-the-loop (HITL) for AI systems. Since its establishment through the DPA, the Office of the Data Protection Commissioner (ODPC) has issued more than 500 determinations or notices to data controllers and processors found to have violated the law and received more than 9,000 complaints,[121] indicating its growing responsiveness and enforcement capacity.[122] The National Data Governance Policy, currently in development, aims to provide Kenya's data ecosystem with a more harmonised legal framework.[123]

118 OECD: "Data governance".
119 Kenya Bureau of Standards (KEBS). Information technology – Artificial Intelligence – Code of Practice for AI Applications.
120 Adams, R. et al. (2025). AI in Africa: A Landscape Study. Global Center on AI Governance.
121 As of January 2019. See: Harrison, M. (26 January 2026). "Over 9,000 Data Privacy Complaints Filed Since 2019, Says ODPC". The Coast.
122 Juma, I. and Faturoti, B. (23 May 2025). "Enforcing data privacy in Kenya and Nigeria: towards an African approach to regulatory practice". International Review of Law, Computers & Technology.
123 Onyeagoro, J. (10 September 2025). "Kenya Begins Drafting National Data Governance Policy." TechAfrica News.

Figure 8: Additional data laws, policies and frameworks in Kenya

Additional data-related regulations, while not explicitly focused on AI, address and protect data rights.	Law	Computer Misuse and Cybercrimes Act (2018); Copyright Act; Access to Information Act
	Draft	ODPC Data Sharing Code
Sector-specific laws touch on data regulation for specific industries.	Law	**Finance**: National Payment System Act (2011) and National Payment System Regulations (2014); **Health**: The Public Health Act (2012), Health Act (2017), HIV and AIDS Prevention and Control Act (2006) and the Digital Health Act (2023)
	Draft	**Agriculture**: Kenya Agricultural Data, Information and Digital Policy[124]
	Framework/strategy	**Geospatial data:** Kenya Earth Observation Data Sharing Framework 2026[125]
Roadmaps and strategies on AI and emerging technologies establish the country's vision and ambitions for AI development.	Framework/strategy	National AI Strategy (2025–2030); AI and Blockchain Taskforce 2018/Distributed Ledgers (Blockchain) and Artificial Intelligence Taskforce Report 2019; Kenya National Digital Master Plan (2022–2032); 10-year National Innovation Masterplan (2023)
National socio-economic development strategies highlight the role of digital infrastructure, technologies and innovation in driving inclusive growth.	Framework/strategy	Kenya Vision 2030; Kenya's Bottom-Up Economic Transformation Agenda
Pan-African frameworks drive more harmonised AI governance and cross-border data exchanges, but are not legally binding.	Framework/strategy	The Continental Artificial Intelligence Strategy (July 2024);[126] African Digital Compact;[127] 2022 AU Data Policy Framework; African Union Convention on Cyber Security and Personal Data Protection (commonly known as the Malabo Convention); AU Digital Transformation Strategy for Africa (2020–2030)

124 Ministry of Agriculture and Livestock Development. (2026). Kenya Agricultural Data, Information and Digital Policy.
125 Moussa, R.B. (30 March 2026). "Kenya Launches National Framework to Govern Earth Observation Data". Space In Africa.
126 Adams, R. et al. (2025). AI in Africa: A Landscape Study. Global Center on AI Governance.
127 African Union. (2024). African Digital Compact.

The Data Protection Act has provided a valuable framework for ethical data sharing, but it comes with some apprehension and uncertainty around data sharing and use. The research showed inconsistent interpretations of the law. The DPA is seen, in turn, as both an enabler of and a deterrent to data sharing and innovation. For some, the punitive measures are a source of fear and a reason to avoid sharing data altogether. Smaller actors in the ecosystem also see it as disproportionately affecting them.

Registering as data handlers and processors for compliance with the DPA has also proved challenging or confusing for some organisations and county governments. These challenges point to the need to expand training and technical support for public and private organisations on the law and compliance.

There is also uncertainty about the benefits and risks of data sharing, as well as the processes involved. In collaborative projects, the question of ownership and sharing can be unclear. **For startups, researchers and development organisations working directly with communities, data sharing is often a value-laden, case-by-case decision**, ultimately coming down to the question: will this benefit the individual or community? These ethical questions do not always fit neatly into standardised processes or frameworks, yet the ambiguity can hinder sharing efforts. With private sector stakeholders, commercial agreements can also be ambiguous.

> "We've found that people use data governance or privacy as a way to stall projects when they're not fully bought into it. Sometimes someone is just scared of giving us data because they don't know how to go about it."
> **– Data and AI company**

In addition to gaps in understanding of data policy in the public and private sector, there is also limited technical understanding of AI and data among policymakers and regulators. This has an impact on how data regulation is applied in practice. The desire of innovators to move fast is often at odds with the regulatory perspective, pointing to a need for better technical translation between different actors and more engagement of industry and communities in policy decisions.

> "We share data for internal use cases, like fraud detection, but the governance structures in place are not mature enough to support data sharing for citizen- and public-facing applications at the moment... We're building an industry AI governance framework for the insurance sector (to address) grey areas concerning things like data retention, AI model selection and cross-border data transfer."
> **– Insurance company**

Licences

Data sovereignty has become an important consideration in Kenya's evolving digital landscape. As the country positions itself as a regional digital hub, efforts are underway to develop frameworks that strengthen data sovereignty, such as the planned establishment of a Trusted Data Zone[128] that would allow foreign organisations to store and process data in Kenya under a regulatory framework aligned with Kenyan laws. A more complex side of data sovereignty is the issue of data extraction by international companies for profit, with little to no benefit for local people, a dynamic likened to digital colonisation.[129] The data demands of large GenAI models only drives these extractive processes, through widespread data scraping and web crawling.[130] For organisations working in public service delivery in particular, the reliance on cloud servers outside the country, as well as the large role played by international partners in Kenya's development ecosystem, raise sovereignty concerns.

Diverse data licensing options and frameworks can help ensure local ownership of data and long-term sustainability of data processes. While open data licences democratise innovation,[131] they can also reinforce global power imbalances, as individuals and organisations in low- and middle-income countries (LMICs) lack the resources to take full advantage of open-source datasets.[132] While intended to foster innovation, it can be challenging for local innovators and researchers to afford to make open data AI ready or sustain data pipelines due to the costs associated with data processes, like anonymisation. Without established mechanisms to compensate or reward local organisations or communities when their data is shared and monetised by larger entities, they are more protective of their datasets. Having multiple, Africa-centric licensing options helps navigate these tensions. For instance, it allows the commercialisation of some aspects of datasets while also keeping them open for other purposes, like research. Licensing models like the Nwulite Obodo Open Data License (NOODL) License[133] and the Esethu License[134] are African alternatives that protect the communities behind the data.[135]

128 Microsoft. (22 May 2024). "Microsoft and G42 announce $1 billion comprehensive digital ecosystem initiative for Kenya".
129 HAI Stanford University. (15 August 2022). "Neema Iyer: Digital Extractivism Africa Mirrors Colonial Practices".
130 Kaaniru, J. (10 April 2025). "Ethical AI Development in Kenya: The Role of Ethical Data Sourcing and Governance". Strathmore University Centre for Intellectual Property and Information Technology Law (CIPIT).
131 Ibid.
132 Licensing African Datasets: Nwulite Obodo Open Data License.
133 Ibid.
134 Shikwambane, N. (16 April 2025). "A Global First: How a New Sustainable Data Framework & License Are Transforming Language AI". Lelapa AI.
135 Strathmore University Centre for Intellectual Property and Information Technology Law (CIPIT). (2025). The State of AI in Africa Report.

Data infrastructure

While the foundation of Kenya's ICT infrastructure is strong, persistent challenges with reliable and affordable internet and connectivity have created gaps in data generation, availability and deployment,[136] and limited the adoption and sustained use of digital data processes in the public sector. For instance, in Turkana County, due to unreliable internet connectivity, a national agency's county-level officials collect data on paper in the field and later digitise it. While digital tools exist for health centres and Community Health Promoters (CHPs), poor network coverage in remote areas makes these tools ineffective or forces them to revert to manual data entry. Murang'a County has taken steps to build more resilient infrastructure and design for their current realities. For instance, the ICT department has implemented offline backup capabilities within local servers that allow data to be captured and stored locally when internet connectivity is unavailable. Once connectivity is restored, the system automatically synchronises with central servers to ensure data is consistent and complete. The county is also adopting solar energy solutions in remote health dispensaries, which keep digital health information systems running during power outages.

Kenya's national AI strategy recognises AI infrastructure development as an important step towards data sovereignty. The Kenya Cloud Policy and DPA encourage public sector agencies to migrate their data and computing services to cloud platforms to improve the efficiency, scalability and resilience of government digital systems. Data centres like Konza Technopolis and Ruaraka Government Data Centre provide cloud-based infrastructure to support digital services and data management for national and county government institutions.

Yet poor network connectivity hinders the transition to cloud data storage for national and county institutions, perpetuating the use of local, internal storage infrastructure. Some more digitally mature counties have adopted cloud storage to improve security and data sharing, financed through local ICT department budgets. Incrementally scaling pilots of AI solutions can help governments determine targeted infrastructure investments.

136 Genesis. (2024). "AI in Africa: The state and needs of the ecosystem".

4. MSME SECTOR

This section assesses and analyses the data ecosystem for the MSME sector and explores the potential AI use cases it can support.

Mapping the data ecosystem

The type of data MSMEs generate, and who the primary data holders are, vary depending on certain features of the business.

Formality and size

Around 28% of MSMEs are licensed by county governments (as of 2016).[137] This leaves the majority of Kenya's enterprises operating outside formal regulatory, banking or analytical systems, and generating data that is fragmented, decentralised and analogue, or stored locally on mobiles, requiring digitisation and integration to make it analytically usable. The formality of a business tends to increase with size, so small and medium enterprises (SMEs) have more regular operations and greater digital adoption, and generate richer, more structured and varied data. Still, most licensed businesses are micro-enterprises given that this type of business outnumbers others in the sector.[138]

Licensed MSMEs are more likely to maintain business records, most commonly in the form of cash sales and receipts.[139] Licensed SMEs also have more detailed and varied records, such as sales and purchase ledgers and bank statements.[140] Most unlicensed businesses do not maintain records of any kind, and those that do typically use personal notes. The lack of business records in the sector is a major data gap.

Table 2: MSMEs by size

Size	Number of employees[141]	Annual turnover[142]	% registered[143]
Micro-enterprises	1–9	Less than KSH 500,000	21.8%
Small enterprises	10–49	KSH 500,000–5 million	66.8%
Medium enterprises	51–250	KSH 5–100 million	72.5%

Industry

According to the MSME Survey (2016), wholesale and retail trade and vehicle repairs account for most licensed and unlicensed businesses. Manufacturing, accommodation and food service industries follow, with around 80% of such businesses unlicensed.[144] Businesses in these industries are diverse and not well organised, and industry-specific data is limited. In contrast, the finance sector has a high percentage of formally registered businesses (57%) and the highest percentage of businesses keeping records (82%), resulting in richer data.[145] While agribusinesses have only 17% of registered businesses and 35% keeping records,[146] the agriculture sector benefits more broadly from a wide variety of farm-, value-chain and macro-level data, given its central role in the Kenyan economy and overlapping importance in the development and public sector.

137 See: Ministry of Co-operatives and Micro-, Small and Medium Enterprises (MSMEs) Development.
138 Ibid.
139 KNBS. (2016). Micro, Small & Medium Establishments: Basic Report 2016.
140 Ibid.
141 As defined by The Micro and Small Enterprise Act of 2012.
142 Ibid.
143 Businesses may have permits from county governments but not be registered with the Registrar of Companies. See: KNBS. (2016). Micro, Small & Medium Establishments: Basic Report 2016.
144 Ibid.
145 Ibid.
146 Ibid.

Digitalisation

Digitalisation of business administration is low for MSMEs in Kenya, limiting the availability of AI-ready transactional data. Extensive use of mobile apps, specialised software and enterprise resource planning (ERP) for business administration is limited, and the rate of digitalisation is even lower outside urban hubs like Nairobi.[147] However, it is worth noting that the COVID-19 pandemic may have boosted MSME digitalisation. For instance, the proportion of online sales increased by 10 percentage points following the outbreak of the pandemic[148] and most businesses reported that their reliance on digital tools had increased since the pandemic.[149] The types of platforms used by MSMEs varies by size, with micro-enterprises tending to use social media applications like WhatsApp and Facebook to advertise and transact rather than formal digital marketplaces.[150]

Data in the MSME sector is distributed across a wide variety of disconnected actors, from MSMEs themselves to government agencies, private sector players and domain organisations representing the various industries MSMEs operate in, each with their own opportunities and barriers.

Table 3: Key data holders and data types in the MSME sector

Statistical data (sector level)

Key data holders	Data sources and features	Barriers
Kenya National Bureau of Statistics (KNBS)	MSME Survey, Economic Survey, Census of Industrial Production, Census of Establishments (baseline sectoral data)	• Out of date (e.g. last national-level MSME survey was conducted in 2016)[151]
Central Bank of Kenya (CBK), KNBS and Financial Sector Deepening (FSD) Kenya	Micro and Small Enterprises Tracker Survey series[152,153] (longitudinal data on enterprise survival, credit access and technology adoption)	• One-off, not regularly updated
Multilateral organisations and researchers (private companies and academia)	For example, the World Bank's Business Pulse Survey during the COVID-19 pandemic	

Statistical data (economy level)

Key data holders	Data sources and features	Barriers
KNBS	Trade and manufacturing statistics (relevant to MSMEs' operating context); Consumer Price Indices (CPI) (gathered through monthly web scraping)	
Kenya Revenue Authority (KRA)	Trade volumes and values	

147 Cruz, M. and Hernandez Uriz, Z. (2022). Entrepreneurship Ecosystems and MSMEs in Kenya: Strengthening Businesses in the Aftermath of the Pandemic. The World Bank.
148 Ferber, T. and Buri, S. (2021). Navigating through COVID-19: A Snapshot on How the Pandemic Affected MSMEs in Kenya. The World Bank.
149 GeoPoll, Africa 118 and African Talent Company. (2023). The Africa MSME Pulse Survey.
150 Caribou. (2020). Digital Behaviors of Kenyan Micro-Entrepreneurs.
151 KNBS. (2016). Small and Medium Enterprises (MSME) Survey 2016.
152 FSD Kenya. (March 2023). FinAccess MSE Tracker Survey.
153 FSD Kenya. (August 2023). FinAccess MSE Tracker Survey – Wave II: Follow-up on 2019 and 2021 FinAccess Household Surveys.

Administrative data (national level)

Key data holders	Data sources and features	Barriers
Micro and Small Enterprises Authority (MSEA)	Business registration data (including income, number of employees, location, sector, ownership structure); audited accounts as part of one-off programmes	• Data is self-reported and can have inaccuracies. • Business registration data is stored locally by the MSEA and on the eCitizen platform. Corrections and updates made during quality checks are not reconciled on eCitizen, leading to inconsistencies.
Government financial inclusion programmes: NYOTA project,[154] Hustler Fund,[155] Women Enterprise Fund,[156] Uwezo Fund,[157] Youth Enterprise Development Fund[158]	Credit uptake, repayment behaviour and borrower demographics for financially excluded MSMEs	• Limited to a small sub-section of businesses
KRA	VAT filings, iTax returns and turnover tax declarations	• Limited to the minority of MSMEs that are formally registered
National Social Security Fund (NSSF) and the Social Health Authority (SHA) or Social Health Insurance Fund (SHIF)	Mandatory payroll compliance data, including employment headcount, formal job creation and workforce stability proxies for SMEs	

Administrative data (county level)

Key data holders	Data sources and features	Barriers
County Finance and Economic Planning departments	Support and budget for MSMEs in County Integrated Development Plans (CIDP), County Fiscal Strategy Papers (CFSP); County Statistical Abstracts (CSA) (indicators on trade and commerce, businesses, prices, agriculture)	
County Revenue Management Systems	Records of business permits, locations, operations	
MSEA county-level offices	Additional MSE monitoring data	• Data collection is manual, stored locally

154 Provides business start-up capital for young people: https://nyotaproject.go.ke/
155 Provides instant loans: https://www.hustlerfund.go.ke/
156 Provides credit to women-owned businesses: https://wef.go.ke/
157 Improves access to finance for youth, women and persons with disabilities for enterprise development: https://www.uwezo.go.ke/
158 Offers financial and business support for youth-owned enterprises: https://www.youthfund.go.ke/

Transactional and behavioural data (business level)

Key data holders	Data sources and features	Barriers
MSME owners	Device-based digital data (data consumption, app usage); digital or physical transaction records	• Low granularity
Mobile money platforms (e.g. M-Pesa)	Daily revenue flows, payments and transaction patterns	• Even when MSMEs use digital tools for payments, cash continues to dominate.[159,160] • Digitally excluded groups, particularly women-owned micro-enterprises, remain underrepresented, making the data incomplete. • Proprietary data is difficult to access.
E-commerce and social commerce platforms (e.g. Jumia, Meta, WhatsApp, Makueni County e-marketing portal)[161]	Structured operational data on order volumes, product listings, sales, customer reviews	
KRA	Customs and cross-border sales data for export-oriented SMEs	

Transactional and behavioural data (financial institution level)

Key data holders	Data sources and features	Barriers
Central Bank of Kenya (CBK)	Survey on MSME Access to Bank Credit[162] (credit flows, approval rates and barriers to access across financial institutions)	
Kenya Bankers Association	MSME credit analysis dashboard[163] (gender-disaggregated data on MSME financing patterns)	
Credit Reference Bureaus	Consolidates historical credit performance, repayment and default records across the formal lending ecosystem	• High default rates and multi-borrowing practices can lead to misleading or ineffective signals
Financial institutions (formal commercial banks and informal groups (chamas) and Savings and Credit Cooperative Organisations (SACCOs))	MSME transaction history, loan history and cash flow	• Tight data-sharing regulations. • Informal financial groups are less likely to have digital or structured data.

Industry data

Key data holders	Data sources and features	Barriers
Industry bodies (Kenya Association of Manufacturers (KAM),[164] Kenya National Chamber of Commerce and Industry (KNCCI),[165] Kenya Private Sector Alliance[166])	Data on members, markets, sector performance	• Limited resources and capacity for data collection and analysis.

159 Cruz, M. and Hernandez Uriz, Z. (2022). Entrepreneurship Ecosystems and MSMEs in Kenya: Strengthening Businesses in the Aftermath of the Pandemic. The World Bank.
160 FSD Kenya. (August 2023). FinAccess MSE Tracker Survey – Wave II: Follow-up on 2019 and 2021 FinAccess Household Surveys.
161 Anderson, Y. (9 July 2020). "The Government of Makueni County of Kenya Launched a Marketplace on CS-Cart to Connect Farmers and Traders with Consumers". CS-Cart.
162 Central Bank of Kenya. (June 2025). 2024 Survey Report on MSME Access to Bank Credit.
163 Kenya Bankers Association's Gender-disaggregated Credit Analysis Dashboard.
164 See: Kenya Association of Manufacturers (KAM)
165 See: Kenya National Chamber of Commerce & Industry
166 See: Kenya Private Sector Alliance (KEPSA)

Key insights

The lack of comprehensive and up-to-date baseline statistical data on the social and economic realities of MSMEs leaves large gaps in our understanding of the sector. Without recent surveys and censuses, the landscape within which MSMEs operate, their needs and their context are all poorly understood, making it difficult to develop tailored digital solutions and evaluate their impact.

While MSMEs generate a large amount of administrative data through licences, permits, taxes, compliance systems and business operations, this data is dispersed across different departments and Kenya's devolved counties, leaving the sector without a single source of truth. For instance, there are multiple parallel and disconnected registration and licensing systems at both national and county levels. **Counties are likely to hold the most accurate and up-to-date data on businesses and their operations,** through local licenses and market fees, which also cover informal businesses, but this data is only partially digitalised and centralised.

Device-based mobile activity data provides some insight into MSMEs' functioning, but the most granular digital data is owned by siloed, proprietary digital platforms. This includes mobile money, e-commerce and social media companies, making the establishment of formal partnerships critical.

Business transaction data and business records could have valuable applications but need to be digitalised before they are AI-ready. MSMEs' largely rely on physical records for business management, particularly micro-enterprises. Organisations like Somo[167] demonstrate how digitalising MSMEs is not just about tools and skills, but also supporting a behavioural and mindset shift.

MSMEs are often suspicious of tax implications and government oversight if they share their data, making more neutral, nongovernmental bodies like nonprofits or industry associations important players in the data ecosystem. Given the tight profit margins at which MSMEs operate, seeing the financial benefit of data sharing or digitalisation is important. Direct, tangible benefits like access to credit increases MSMEs' willingness to share their data and foster trust in digital solutions.

> "Nano businesses associate digitisation with formalisation, being in the government's radar, and paying taxes. We need to alleviate that fear. The larger they [businesses] grow, the more they appreciate digitisation, they realise things work much better."
> **– Nonprofit in MSME sector**

167 See: Somo

AI use cases for MSMEs

Despite being central to Kenya's national economy, the MSME sector faces persistent unmet needs. Kenya's tech ecosystem has seen several digital innovations tailored to MSMEs, particularly to address the credit gap – a key barrier for micro-entrepreneurs in particular, who are more likely to be women.[168] AI has already begun to enhance these solutions, such as AI-powered credit scoring and alternative data by startups such as Numida,[169] JUMO,[170] Sokohela[171] and Pezesha,[172] and digital lending institutions like Tala[173] and Branch.[174] However, the datasets used in these models are often black boxes, not openly available to the public.

AI-enabled tools can also support MSMEs by providing financial and business management,[175] e-commerce and customer engagement[176] and supply chain management.[177] Fastagger's edge AI solution translates a business's financial data into actionable insights using small language models (SLMs).[178] To expand its reach, it partnered with Safaricom to be embedded as a "mini app" within the M-Pesa app, highlighting the important role of MNOs in the distribution of AI-enabled services.

There were several use cases in the MSME sector in which the analytical, predictive and generative capabilities of AI could have positive impacts. Two use cases are analysed and mapped here, selected based on the added value of AI, data feasibility and technological readiness (e.g. existence of pilots and deployments). Additional use cases can be found in Appendix 2.

168 CGAP. (2025). Diverse Paths: Finance for Women's Nano and Micro Enterprises.
169 Numida provides digital financial services like credit, savings and insurance to micro- and small enterprises in Kenya and Uganda. See: https://ke.numida.com/about-us
170 JUMO is a financial technology company that partners with banks, MNOs and other e-commerce players to deliver progressive financial choices to customers in emerging markets across Africa and Asia. See: https://jumo.world/
171 Sokohela provides digital loan products to SMEs in Kenya. See: https://sokohela.com/
172 Pezesha, headquartered in Kenya, built a digital financial infrastructure that connects SMEs to working capital. See: https://pezesha.com/
173 See: TALA
174 See: Branch
175 For instance, see: Lipabiz
176 For instance, see Tappi
177 For instance, see Wasoko
178 See: Fastagger

01

Alternative credit scoring to improve access to finance

Need

Access to finance is a major barrier for MSMEs, particularly micro-entrepreneurs and women, who are more likely to lack the documentation and collateral needed for bank loan approvals.[179] Without capital to invest in their businesses, the competitiveness, productivity and growth of MSMEs are all constrained.

AI value proposition

AI prediction models can assess the creditworthiness of unbanked or "thin file" micro-enterprises more effectively, by finding patterns across traditional structured data (e.g. financial records) and unstructured, alternative data (e.g. behavioural data) to classify creditworthiness, predict the likelihood of loan repayment and de-risk lending. Any digital solution needs to be combined with efforts to fix the underlying structural barriers and financial management challenges that contribute to financing gaps for MSMEs.

Table 4: Alternative credit scoring relevant data holders and datasets

MSME transactional data	• Business administrative and management data (sales, stock and customer records)
	• Mobile device-based data (SMS logs, activity history, app and data usage)
Digital platform behavioural data	• From platforms like social media, e-commerce, mobile money platforms
Financial organisation data	• Formal financial institutions (commercial banks, cross-lender credit performance data housed by CRB)
	• Informal financial organisations (informal savings groups (chamas) and small-scale SACCOs)
	• Users' own financial records and statements
Government registration and programme data	• From public financial inclusion initiatives like the Hustler Fund, Women Enterprise Fund and Uwezo Fund

179 UN-Women. (2018). Background Paper: Driving Gender-Responsive Financial Inclusion Models in Africa.

Figure 9: Alternative credit scoring data assessment

Data generation	Smaller digital data footprints of informal businesses, women and rural MSMEs risks discriminatory outcomes
Data quality and reliability	Incorrect or incomplete data can lead to inaccurate creditworthiness assessments
	Behavioural factors, like borrowers taking out multiple loans from different digital lending platforms, can corrupt the predictive signals needed for accurate scoring models
	Limited evidence of how some alternative datasets (input) are linked to creditworthiness (output)
Data integration and interoperability	Poor interoperability and standardisation of diverse datasets
	Limited access to platform proprietary data restricts a holistic borrower view
	Users' device-generated data lacks granularity and can result in false signals
Governance	Private and sensitive data is vulnerable to breaches or exploitation (e.g. used by lenders to harass borrowers for repayment) without sufficient protection
	ODPC has enforced compliance with the DPA for online digital lenders through penalty notices, audits
User and use case alignment	Strong incentive alignment (access to credit) can drive MSME digitalisation

High concern Medium concern Opportunity

02

Business analytics for MSMEs

Need

Kenya's MSMEs have low managerial, financial, technical, technological, entrepreneurial and industry-relevant skills.[180] While training opportunities exist, many MSMEs are either unaware of them or unable to access or afford them, and formal training programmes are unable to address the emerging issues businesses face.[181] MSMEs also lack the knowledge to navigate complex regulatory, legislative and bureaucratic processes.[182] These challenges can all hinder business growth, survival and customer retention.

AI value proposition

For MSMEs with low digital literacy that are unable to use traditional data analytics tools, conversational AI and SLMs can provide access to personalised business insights. SLMs can provide tailored advisory that translates raw, unstructured, context-specific financial and sectoral data into natural language conversations. They can also help streamline fragmented regulatory requirements to aid the business registration process and accurately calculate demand forecasting by aggregating market and behavioural signals that are impossible for micro-enterprises to calculate manually.

Table 5: Business analytics relevant data holders and datasets

MSME transactional data	• Business administrative and management data (sales, stock and customer records)
	• Mobile device-based data (SMS logs, activity history, app and data usage)
E-commerce and social commerce platform data	• Social media, e-commerce, platforms where MSMEs operate and advertise
Market intelligence data (to forecast demand, predict market scarcity, adjust pricing strategies)	• KNBS Consumer Price Index (CPI)
	• Industry bodies
Business registration and licensing data	• National government (MSEA registry on the eCitizen platform)
	• County governments

180 See: Ministry of Co-operatives and Micro-, Small and Medium Enterprises (MSMEs) Development.
181 Ibid.
182 Hinga, J. (18 June 2024). "Unlocking Kenya's Economic Potential through MSME Development". Strathmore University Business School (SBS).

Figure 10: Business analytics data assessment

Data generation and digitalisation	High concern	Less digitalised businesses have a smaller digital data footprint and less granular transaction data than is needed to customise AI outputs, resulting in less accurate model outputs
Quality and reliability	High concern	Market statistics and administrative updates have fluctuations and lags
Integration and interoperability	High concern	Proprietary silos and weak linkages between national- and county-level data
User and use case readiness	Medium concern	Even though conversational AI can simplify business analytics, MSMEs still need to have the skills and trust to use it, making accuracy, reliability and user feedback loops important
Infrastructure	Opportunity	Using SLMs and Edge AI to process data entirely on MSMEs' devices can address both infrastructure and privacy barriers

High concern Medium concern Opportunity

Summary

Digital inclusion and access barriers affect data readiness and, ultimately, the effectiveness of AI solutions. Informal, women-owned and rural enterprises are all more likely to have a smaller digital data footprint and less granular data to support customised AI outputs, resulting in models that are less accurate, or even discriminatory, for these already marginalised groups. Barriers accessing proprietary datasets (e.g. from financial institutions and e-commerce platforms) adds to the reliance on MSMEs' device-generated data as proxies, but this data can result in false signals. For instance, a user's M-Pesa statement might not distinguish between personal and business transactions with the level of granularity needed. Strong private sector partnerships, alongside compliance with data regulations, are key.

The risks of inaccurate, unethical and discriminatory outcomes are high. For the credit-scoring use case, for instance, an MSME may be wrongly denied a loan, or conversely, a loan that is calculated too high can present the risk of defaulting in a sector where default rates are already high. When it comes to business analytics, incorrect advisory can affect MSMEs' business decisions and, ultimately, the livelihoods of business owners. This calls for strong monitoring of bias in different demographics and human oversight across the AI value chain. It also points to the need for inclusive data practices, like gender-disaggregated data, to promote accessible and equitable solutions. While AI has expanded the use of non-traditional or alternative data, the stability and reliability of new types of data remains to be seen.[183] Ensuring transparency in data used and explainability of AI models is necessary to build the trust of MSMEs and the accountability of innovators.

183 The World Bank. (2024). The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges.

5. PUBLIC SERVICE DELIVERY SECTOR

This section assesses and analyses the data ecosystem for the public service delivery sector and explores the potential AI use cases it can support.

Mapping the data ecosystem

Data relevant to public service delivery is generated across three main streams: administrative data from sector-specific platforms, statistical data from nationally coordinated surveys and programmatic data from development partners and NGOs operating parallel to government systems.

Public service delivery infrastructure is centralised within Kenya's eCitizen platform, which hosts more than 21,000 digital government services and captures operational data detailing service access, timing, location and channel.[184] In combination with physical Huduma Centres, eCitizen constitutes a unified, cross-sectoral, citizen-generated administrative data layer. Additional data pillars are often sector-specific, with data maturity varying by type of public service, as detailed in Table 6.

Table 6: Key data holders and data types in the public service delivery

Health data

Key data holders	Data sources and features	Barriers
Ministry of Health	Kenya Health Information System (KHIS), the national DHIS2 platform (routine facility-level data across all counties); Kanya Master Health Facility List; Integrated Human Resource Information System (health workforce and human resources); Community Health Information System (eCHIS) (community health services, supplies, household and patient data)	The Digital Health (Data Exchange Component) Regulations (2025) established a national exchange mechanism and enforced interoperability standards, but implementation is limited by a lack of awareness and compliance costs. Interoperability often still requires manual data extraction and re-entry. eCHIS: Lacks mechanisms to validate entries or track changes, which can compromise data accuracy.[185]
County governments	Electronic medical records (EMRs); medical equipment and supplies; health budgeting	
KNBS	Kenya Demographic and Health Survey (population-level health estimates)	Not regularly conducted
Social Health Authority (SHA)	TaifaCare (patient-level longitudinal data)	
NGOs and development partners[186]	Health programmes and patient-level data	Sits outside the national public health architecture and is not interoperable.
Private health facilities	Health records and claims	
Safaricom M-TIBA (mobile health wallet linked to M-Pesa)	Health payments, spending, costs	
Insurance companies	Digital health records	Sits outside the national public health architecture
Research institutes and universities (e.g. The Kenya Medical Research Institute, The University of Nairobi's College of Health Sciences, Moi University School of Medicine)	Local health research	

184 Africa Check. (June 2025). eCitizen Bulletin, Vol. 1, Issue 5.
185 Amref. (15 October 2025). "Harnessing Generative AI to Transform Community Health: Lessons from Machakos County".
186 For example, programmes such as PEPFAR hold their data in DATIM, a PEPFAR-specific version of DHIS2.

Education data

Key data holders	Data sources and features	Barriers
Ministry of Education	Kenya Education Management Information System (KEMIS) (assigns every student a unique personal identifier (UPI) to track records across all levels)	
Kenya National Examinations Council	Student outcomes	External access is tightly controlled
Teachers Service Commission (TSC)	Administrative data on teacher staffing, payroll	
Kenya Universities and Colleges Central Placement Service (KUCCPS)	Administrative data on university placements	
Private sector digital learning platforms, educational institutions	Student behavioural patterns, student records	Although the private sector is heavily involved in education, data from private institutions is often missing from data collection and analysis.[187]

Population registration data

Key data holders	Data sources and features	Barriers
Ministry of Interior and National Administration and the National Registration Bureau (NRB)[188]	Integrated Population Registration System (IPRS)[189] (centralised demographic and biometric records of citizens)	
Ministry of Lands and Phyiscal Planning	Ardhisasa platform[190] (national land transaction and ownership records)	
Directorate of Civil Registration Services (CRS)	Continuous birth and death registrations, incorporating the Maisha Namba as a lifetime UPI	
State Department for Social Protection	Enhanced Single Registry (ESR) (socio-economic database of vulnerable households and cash transfer recipients)	National dataset disconnected from independent, county-level welfare initiatives
County governments	Localised operational data	
NGOs and implementing partners	Programmatic data	Disconnected from government systems, project-based and time-bound, can be duplicative

187 Open Institute. (2022). Kenya Counties Data Maturity Survey Report.
188 See: https://nrb.ecitizen.go.ke/
189 See: Integrated Population Registration System (IPRS)
190 See: Ardhisasa

Agricultural data

Key data holders	Data sources and features	Barriers
Agriculture and Food Authority (AFA)	Production, licensing and export market data across regulated value chains	
Researchers, local organisations, development partners (e.g. LDRI, vuli.ai[191])	Farm-level, near real-time data on crop health, value chains, financing, agricultural risk modelling	Data access barriers
Agri-techs (e.g. Shamba Records, DigiFarm)		
Ministry of Agriculture and Livestock Development (and KNBS)	National Agriculture Production Report (annual verified, county-level dataset on crop area, production volumes and value	
Ministry of Agriculture and Livestock Development	KilimoSTAT (open data platform for the sector); Kenya Integrated Agriculture Management Information System (KIAMIS) (farmer and agricultural services information)	
	Geospatial data, as listed in Section 1	

191 See: vuli.ai

Key insights

Major progress is being made to improve data digitalisation and interoperability at a national level, but Kenya's devolved governance structure poses an inherent obstacle. Data held at the county level is disconnected not only from other counties, but also the national government. There is also a disparity in the reliability of data between national sources like KIAMIS, which are verified by external auditors, and local datasets, which are not. This has implications for the level at which AI-based public service delivery interventions are designed.

Counties are at different stages in their digitalisation journey. While some offices have locally digitalised data, many are still paper based. Even within a county, different departments have different degrees of digitalisation. Socio-economic contexts also have an impact on data practices. For instance, in the context of agricultural data, counties with a large portion of farms organised into cooperatives may have more consolidated data. Solutions therefore need to be tailored to each county.

The heavy presence of development partners and the private sector in areas like health, education and agriculture enriches and complements data collection efforts. For instance, NGOs working in the health sector, for instance, supplement health system data, and health departments work closely with them, regularly using their data for coordination, planning and operations.[192] At the same time, these datasets sit outside existing systems, contributing to fragmentation, and project-based data collection efforts are often short term.

Data silos prevent the development of comprehensive datasets, which are needed to build solutions for the complex, inherently cross-sectoral challenges in public service delivery. Data strengthening efforts happen within sectors rather than across them, and bridging these efforts is a challenge in the absence of strong DPI and interoperable ID frameworks. For instance, an AI-powered initiative for nutrition required bringing together health, agriculture and satellite data.[193] Assistive Technology for Disability Trust (AT4D)'s work to link children with disabilities to public services faced barriers due to a disconnect between different ministries, such as health, education and social protection, preventing the coordination and flow of data needed to build effective life-course solutions.

Establishing the trust of citizens and frontline workers in data practices, data protection and the AI systems built on top of their data is particularly important in the public sector. Failed digitalisation initiatives, combined with data vulnerabilities related to digital identification and biometric data,[194] have produced some public mistrust and scepticism of government digitalisation efforts.[195] This will need to be managed with caution when introducing AI-based systems in the public sector, as there is the risk of eroding citizens' trust in public services. Digitalisation in the public sector also runs the risk of entrenching issues like corruption and mismanagement.[196]

192 Open Institute. (2022). Kenya Counties Data Maturity Survey Report.
193 Microsoft. (10 October 2024). "Amref Health Africa predicts and prevents malnutrition in Kenya with AFI".
194 Alayande, A., Segun, S. and Junck, L. (2025). Emerging technology policies and democracy in Africa. The Atlantic Council.
195 Zollmann, J., Sambuli, N. and Wanjala, C. (August 2024). Citizen Experiences with DPI: Kenya's Digital ID Transition.
196 Strathmore University Centre for Intellectual Property and Information Technology Law (CIPIT). (2024). An In-depth Analysis of the AU AI Continental Strategy and Implications on AI Governance in the Continent.

AI use cases for public service delivery

The adoption of AI in Kenya's public sector is at an early stage. The AU's Continental AI Strategy highlighted the promising role of AI in improving service delivery, including tax collection and benefit payments. Machine learning and AI-powered predictive models can support analysis of public data and forecasting capabilities, allowing policymakers to make more evidence-based public service delivery decisions, contributing to efficiency and effectiveness. AI can also be used by governments to automate routine processes, detect anomalies for greater accountability and better understand, engage with and tailor services to citizens' needs, improving the responsiveness of service delivery.

Already, the Kenyan government has stated that it is integrating AI in public service delivery,[197] has developed AI skill-building initiatives for public officials[198,199] and built AI-enabled chatbots to help citizens access government services[200] and comply with data protection requirements.[201] The Kenyan judiciary is also exploring the use of AI for internal operations – a prevalent AI use case globally (relative to other government functions)[202] despite the high risk associated. This could be due to both the huge volume of need and backlog and the prevalence of comprehensive and structured data in justice systems, underscoring the role of data availability in prioritising use cases.

Two high potential use cases have been analysed and mapped here, with additional use cases in Appendix 2.

197 State Department for Public Service. (2025). Kenya charts path toward AI-driven public service.
198 Joint SDG Fund. (2025). "Kenya launches pioneering AI and digital courses for public servants".
199 See: Africa Center of Competence for Digital and Artificial Intelligence Skilling.
200 See: EGOV1 – AI Chatbot for Discoverability of Government Services.
201 The Judiciary: "New Technology to Automate Court Processes".
202 OECD. (18 September 2025). Governing with Artificial Intelligence.

01

Agricultural subsidy disbursement

Need

Agricultural input subsidies can help address Kenya's challenges with food insecurity, food price volatility and low crop yields, especially in arid and semi-arid regions.[203] The National Fertilizer Subsidy Programme (NFSP), for instance, is a big part of Kenya's agriculture policy, but is undermined by challenges like inadequate attention to soil health, limited integration with broader agricultural support systems and a failure to tailor inputs to local agro-ecological needs and conditions.[204,205] Efforts like Africa Fertilizer Watch and Kenya Fertilizer dashboard,[206] that track fertilizer price trends and market risks at a regional and national level, have championed more data-driven decision-making at a policy level.

AI value proposition

AI models can be used to identify farm-level vulnerabilities to diseases, pests and nutrient deficiencies; direct targeted and tailored distribution of subsidised inputs by national and county governments; and integrate other agricultural data like soil health, soil testing and weather data. AI can structure fragmented biodata, production data and economic data collected from field officers to build verifiable farmer profiles using for targeted distribution of subsidised seed and fertilizer.

203 Boke Rioba, N. (2025). "Re-Assessing Agricultural Subsidies in Kenya: Successes, Challenges and Opportunities for Enhancing Food Security". Scholars Journal of Agriculture and Veterinary Services.
204 Ibid.
205 Ayalew, H. et al. (6 June 2025). "Rethinking Fertilizer Subsidies in Kenya Towards More Inclusive and Sustainable Models". CGIAR.
206 See: AfricaFertilizer

Table 7: Agricultural subsidy disbursement relevant data holders and datasets

Geospatial data maps for farms	• High-resolution satellite imagery (Planet, Sentinel, Modis) mapped against GPS demarcations of smallholder farm boundaries to track land use and identify localised vulnerabilities to crop failure, pests and nutrient deficiencies.
Verified agronomic knowledge bases	• Crop calendars and disease advisory guidelines, real-world farmer Q&A pairs (from research institutions)
IoT sensor data	• To track factors like soil pH
National farmer registries	• For targeted distribution of subsidised seed and fertilizer.
Locally generated farm reports	• Such as county agricultural extension officer reports and individual farmer records in farmer booklets
Parastatal data	• Production, storage and export metrics from government parastatals, such as the National Cereals and Produce Board, Agriculture and Food Authority (AFA), local cooperatives

Figure 11: Agricultural subsidy disbursement data assessment

Data generation and digitalisation	Hyper-local extension worker and farmer booklet data remains paper-based and not machine readable
	High-resolution geospatial and satellite imagery for farm boundaries is expensive and often performed only periodically due to expense
Quality and reliability	National records are generally accurate, complete, verified and audited
	Locally generated reports are often unverified and quality varies depending on the skills of frontline staff, leading to possible errors and data gaps
Integration and interoperability	Data is siloed across counties and parastatals
User and use case readiness	Farmers and frontline extension staff can be sceptical of, or fatigued from, data collection and sharing efforts

High concern   Medium concern   Opportunity

02

Caseload triage for community health promoters

Need

Kenya's Ministry of Health has expressed interest in deploying AI to support its network of more than 100,000 community health workers, or community health promoters (CHPs), who help provide primary healthcare and universal health coverage through household visits and monitoring, immunisation support, health education, data collection and referrals.[207,208] In a sector where timeliness is critical, CHPs use basic digital tools that can create a delay between data collection and decision-making.[209] CHPs are also burdened with time-consuming administrative tasks.[210]

AI value proposition

GenAI can help CHPs optimise their routes and prioritise high-risk cases by analysing patient records, location, case urgency and broader disease trends and patterns.[211] These AI use cases must work hand in hand with material and systemic support for CHPs. CHPs are not fully integrated in the health system and often excluded from important decision-making.[212] AI solutions need to counter this exclusion rather than enhance it.

Table 8: Caseload triage for CHPs relevant data holders and datasets

Clinical and disease-tracking data	• Health information systems (HIS) databases and patient records like e-CHIS, DHIS2, Taifa Care
Demographic surveys	• Primary household and census data, like the KNBS Demographic and Health Survey (baselines to accurately profile at-risk households)
Geospatial data	• In order to map the most efficient routes for CHPs
Nonprofit and research institution data	• Covering health trends and population-level data

207 iAfrica. (6 March 2026). "Kenya to Deploy AI in Community Health System to Support Frontline Workers".
208 Amref. (17 June 2025). "Built from the Ground Up: How 107,000 Community Health Promoters are Changing the Face of Health Care in Kenya".
209 Ibid.
210 Vota, W. (12 February 2026). "What Are Generative AI Solutions for Community Health Workers?" ICTWorks.
211 Amref. (15 October 2025). "Harnessing Generative AI to Transform Community Health: Lessons from Machakos County".
212 Mwenda, M. (n.d.). "The Role of Community Health Workers in East Africa: Insights, Challenges and the Path Forward". CHW Central.

Figure 12: Caseload triage for CHPs data assessment

Data generation and digitalisation	Opportunity	Routine generation of digital clinical and disease-tracking data through mandatory patient records at health facilities
Quality and reliability	High concern	Clinical tracking systems like Taifa may contain data gaps and inaccuracies that affect the accuracy of AI model outputs
	High concern	Lags in national systems like DHIS2, such as a delay between an incident and its entry in the system, can compromise the timeliness of data for decision-making
Integration and interoperability	High concern	The absence of universal unique patient identifiers (UPIs) makes data fragmented (although the Health Sector Unique Patient Identification framework (2022) is being operationalised to address this)
Governance	High concern	Privacy concerns around data sharing are heightened in the health sector, making strong governance systems when sharing health data essential.
User and use case alignment	Medium concern	CHPs' buy-in and trust must be built to ensure AI solutions do not add to operational demands

High concern   Medium concern   Opportunity

Summary

These two use cases demonstrate the variations within public service delivery. The healthcare sector routinely collects high volumes of operational data and has benefited from strong national implementation structures and centralised databases for data management.[213] Agricultural data, on the other hand, is harder to standardise, as both agricultural activities and data processes vary by county and across different value chains, without a single top-down data management system.[214] Different public services involve different sets of actors and contexts, indicating the need for tailored approaches. Paper-based, manual, ad hoc and non-standardised data processes make it difficult to unify data systems across counties and agencies, indicating that smaller, tailored solutions may be more appropriate.

The use cases also demonstrate that trust is the bedrock of data collection and adoption of digital solutions. Farmers, frontline extension staff, CHPs and community members all need to be able to see the benefits of data sharing and data collection while tackling the fatigue, scepticism or apprehensions that can come with it. In addition to being end users, CHPs are also key data collectors in Kenya's health system, and it is important to ensure that AI applications do not add to their operational demands. Generating their buy-in and trust across the AI value chain is essential.

Inaccurate data in the public sector can have serious consequences: in the CHP use case, false negatives and false positives can deepen health inequalities in the marginalised communities that CHPs typically work with,[215] while inaccurate subsidy disbursement has a direct impact on the operations and livelihoods of farmers. A recent investigation into the use of a machine learning model by the Social Health Authority (SHA) to predict households' income, and thus determine Kenyans' health insurance contribution, revealed the major real-world harms that AI in public service delivery can have.[216] Outdated, non-representative and flawed training data and opaque algorithmic decisions impacted vulnerable Kenyans' financial security and health access, as the model incorrectly predicted household wealth, overcharging the poorest households for health insurance. With such a high potential for harm, rigorous data verification, quality control, risk assessment and model testing, transparency into AI-based predictions and decisions, and the right to contest automated predictions or decisions are all essential.

213 Open Institute. (2022). Kenya Counties Data Maturity Survey Report.
214 Ibid.
215 Njoroge, I. (June 2025). "You are just like them: The paradoxical position of Nairobi's community health promoters, a photovoice study". SSM – Health Systems, 4.
216 Mukami, P. et al. (May 2026). "Hiding Behind AI: How SHA Was Used to Load Health System Costs Onto Poorest". Lighthouse Reports.

6. RECOMMENDATIONS

There are a variety of promising efforts to strengthen Kenya's data ecosystem. The API Gateway Strategy and GAVA Connect (currently used by the KRA) are bridging the gaps between fragmented government data systems and support open APIs to link internal data and counter data silos. At a governance level, the ODPC has sector-specific guidelines[217] that could be expanded, and the regulatory sandboxes for Kenya's financial sector[218] are creating an opportunity for regulators and innovators to work together in real time. Development partners, civil society and NGOs are building, funding and supporting a number of data initiatives.

Based on the barriers identified in the research, the following recommendations can add to a more effective and collaborative data ecosystem capable of supporting AI solutions.

National and county governments

1. Data digitalisation and standardisation:	Uphold and support the implementation of standardised guidelines for public sector data collection and data digitalisation, definitions of key metrics and metadata formats.	Medium term
2. Data sharing:	Standardise data-sharing MoUs in the public sector to move from ad hoc, individual agreements to a common framework and process for commercial and research data valuation and exchange.	Long term
3. Cross-sector coordination:	Establish mechanisms such as a coordination council of public and private sector experts to ensure that sectors can speak to each other, understand overlaps and redundancies in their data collection efforts, recognise shared value and implement data-sharing agreements. A coordination council can also focus on priority "golden datasets" for high-impact, cross-sectoral tasks, such as parallel translation into local languages or localised crop disease imagery that can be treated as digital public goods.	Long term
4. County-level data practices:	Establish pathways for counties to share real-world experiences, lessons and blueprints of successful digitalisation efforts and AI deployments, potentially through county sector working groups within the Council of Governors.	Short term
	Expand efforts to establish data expertise within counties. Consider appointing county "data champions", their role formally defined and institutionalised, to ensure data quality at the point of entry and maintain historical data records during government transitions. This expertise can include statisticians in addition to data scientists as a more readily available skillset.	Medium term

217 See: ODPC Guidelines
218 Pathways for Prosperity Commission. (2019). "Kenya: A regulatory sandbox for the financial sector". Case study 9. Digital Economy Kit.

5. Data collector engagement:	Bring frontline workers into the decision-making process when developing new data practices and systems. As part of frontline worker training, highlight how ground-level data ultimately informs decision-making.	Short term
6. Address vendor lock-in:	All public sector IT contracts should mandate escrow agreements and interoperability in procurement, ensuring databases and source code belong to the government upon contract termination.	Long term
7. Shared national cyber infrastructure:	Consider investing in high-performance computing (HPC) and cyber infrastructure under one umbrella, enabling shared access to compute and data infrastructure, reducing duplication and supporting innovation across academia, government and industry. This is similar to South Africa's National Integrated Cyber Infrastructure System (NICIS), which provides a unified structure for governing essential digital resources, preventing the fragmentation and duplication that currently hinder Kenya's AI ecosystem. At a government level, the compute divide can be countered by supporting and building shared sovereign infrastructure, such as the Council of Governors' initiative to create shared resources and infrastructure for counties.	Long term

Other ecosystem players

Private sector

1. Demonstrate value:	Show what can be done with existing public sector datasets, even when they are limited. Build and document proofs of concept and pilots to highlight the added value of data and secure buy-in and investment in stronger data processes.	Short term
2. Innovator collaboration:	Foster collaboration between startups on foundational data efforts (e.g. digital data pipelines), which can have multiple applications, to save resources, avoid duplication and distribute risk.	Medium term
3. Data sharing frameworks:	Establish data frameworks that support innovation while maintaining commercial interests. For example, limiting access to certain datasets or geographies, developing structured access models (e.g. metered or tiered access) and ensuring pilots convert into ongoing, revenue-generating relationships.	Long term

Funders, development partners and civil society

1. Capacity building and sustainability:	Invest in the lasting capacity of public sector organisations over one-off, project-based systems. Align data collection processes and systems with local realities; co-develop with local government IT teams; plan for the long-term sustainability and ongoing maintenance of data systems.	Short term Medium term
2. Licensing:	Fund fair compensation for data subjects and allow for context-specific, benefit-sharing licensing frameworks.	Short term
3. Ecosystem convening for data access:	Support structured access to data using convening power. For instance, the World Bank's Soil Health Innovation Challenge made more than 19 anonymised datasets from various partners available to innovators in Kenya, and ensured innovators could retain access to datasets even after the challenge.[219] The World Bank incentivised data holders by offering them the opportunity to develop use cases and business models for their data.	Short term Medium term
4. Data scorecards:	Publicly track and audit public sector datasets to identify deployment, opportunities, gaps, privacy and inclusion.	Short term
5. Data rights:	Expand efforts to build awareness and understanding of data rights and data governance at a community level and support community-centred benefit sharing and licensing.	Short term

219 World Bank's Data for soil health and innovation challenges 2025

APPENDIX

APPENDIX 1: LIST OF EXPERTS CONSULTED

Table: Breakdown of stakeholders and number of organisations consulted

Stakeholder type	Number of organisations consulted
County-level public sector stakeholders from different geographic regions and contexts with varying data-readiness levels	3
National-level public sector stakeholders across relevant ministries and departments	5
Private sector stakeholders, such as innovators, startups and organisations/companies with domain-relevant data	13
Domain experts, such as development actors, researchers and academics, NGOs and community groups relevant to the focus sectors	20
Global experts working on data and AI in comparable markets	2
	Total: 43

We would like to thank all the organisations and individuals that enriched this study by sharing their insights. Some of them are listed here.

Advisors:
Linet Kwamboka and Leonida Mutuku

Organisations that contributed to the research:

- Aapti Institute
- AI Kenya
- Assistive Tech 4 Development (AT4D)
- @iLabAfrica
- Dalberg Data Insights
- Data Labelers Association (DLA)
- Digital Umuganda
- Fastagger
- Humanitarian OpenStreetMap Team
- ICON Data and Learning Labs (IDL)
- Jacaranda Health
- JHUB Africa
- Kenya National Chamber of Commerce and Industry (KNCCI)
- Kilifi County
- Micro and Small Enterprises Authority (MSEA)
- Murang'a County
- Numida
- Pathways Technologies Limited
- Power Learn Project
- Prescott Data
- Qhala
- Shamba Records
- Ushahidi
- UzimaNexus

Individuals who provided guidance and/or attended the roundtable gathering in Nairobi:

- Amos Mwendwa
- Charity Wayua
- Christine Ajulu, Open Institute
- Emmy Chirchir, IDRC
- Girmaw Abebe Tadesse, Microsoft AI for Good Lab
- Isaac Looremeta, Numida
- Jessica Wangai, Action Lab
- Michael Michie, Everse Technology
- Nasubo Ongoma, Qhala
- Nzau Muinde, GSMA
- Dr. Paul Korir, Center for Epidemiological Modelling and Analysis (CEMA), University of Nairobi
- Stella Gichuhi-Karau, Everse Technology
- Winny Chelangat, Datawise Africa

APPENDIX 2: ADDITIONAL USE CASES

i. MSMEs: Supply chain optimisation

Impact potential: Optimising routes, matching supply with demand and tracking goods from the farmgate to the export market directly affects MSMEs' profitability and market access.

Added value of AI: AI could support supply forecasting, production planning and optimisation of delivery routes – complex tasks that are nearly impossible for micro-enterprises to manage manually.

Data sources: Alternative transactional data (e.g. M-Pesa statements); e-commerce interaction and fulfilment data; geospatial and GPS routing data (real-time location data collected directly from drivers' mobile phones or vehicle trackers); public transport and traffic data (from smart city infrastructure and government administrative records)

Feasibility: While there are accurate and consistent location and GPS data for logistics analysis, the logistics sector faces broad operational and environmental challenges. Financial and market data are often incomplete, while inconsistencies in pricing data over time introduce noise to predictive models. Data is also heavily siloed across different transport, logistics and e-commerce platforms, making it difficult to integrate.

ii. MSME and public service delivery: hyperlocal agricultural advisory and yield forecasting

Impact potential: Providing hyperlocal, AI-driven early warnings for crop failure, pests and drought has a direct impact on food security and climate resilience for highly vulnerable smallholder farmers and agribusinesses in Kenya.

Added value of AI: AI offers capabilities that cannot be achieved manually at scale. Computer vision models can instantly diagnose nutrient deficiencies or pest damage from smartphone images, while predictive ML models can fuse satellite imagery with ground-truth data to forecast yields and optimise fertiliser use. Additionally, SLMs and NLP can translate complex agronomic advice into local dialects and conversational, voice-based interfaces making advisory more accessible. County government interviewees also pointed to the value in offering advise to farmers on what and when to plant, based on analysis of crop yields, weather, and crop price variations.

Data sources: Geospatial and satellite data; primary farm-level data; computer vision data (images of crops that have been manually annotated by agronomists); real-world Q&A pairs (authentic, culturally relevant questions and answers sourced from actual end users); language and conversational data; farm-level sensor and IoT data; secondary institutional data (e.g. crop calendars, indigenous vegetable data and historical production records)

Feasibility: High-resolution satellite imagery is prohibitively expensive for local NGOs and startups, and agricultural data in Kenya is siloed across government agencies, counties and private platforms. There is also a shortage of annotated local language audio data and a lack of real-world farmer Q&A pairs required to train conversational advisory bots and to test how well AI models provide culturally relevant, localised advisory.

iii. MSME and public service delivery: land use mapping for agricultural MSMEs

Impact potential: Accurate land use and land cover maps allow authorities to estimate actual agricultural production capacity based on the specific crops being planted. This directly informs critical public sector decisions, such as supporting farmers, setting accurate market prices, managing food security and planning for agricultural expansions. It also supports environmental conservation by tracking land degradation over time.

Added value of AI: Manually mapping vast tracts of agricultural land is nearly impossible. Geospatial AI (GeoAI), which applies computer vision and ML to satellite or drone imagery, offers exponential improvements by automatically identifying land features, crop types and boundaries at scale.

Data sources: Geospatial and satellite imagery; high-resolution drone imagery; ground truth data (real-time data collected directly from the field to validate satellite images, including exact farm boundaries (geotagging) and crop varieties planted); labelled map data (to train the models to accurately identify buildings, roads, waterways and agricultural plots)

Feasibility: High-resolution satellite imagery is prohibitively expensive for smaller actors. Collecting localised aerial data via drones is hindered by unclear, non-specific regulations on drone flights in Kenya. Lastly, processing massive geospatial datasets requires high-performance compute infrastructure that is often too costly for local innovators.

iv. Public service delivery: citizen engagement

Impact potential: Citizen engagement can help make public services more transparent, accountable and trusted, yet remains hard to incorporate in public service decisions at scale.

Added value of AI: Social media and crowdsourced channels generate millions of unstructured data points daily, making manual processing impossible. AI models could scrape and analyse social media platforms like Facebook, X (Twitter) and TikTok to extract real-time citizen perspectives, outside of formal town hall meetings, as well as to identify signals like misinformation outbreaks or urgent disaster alerts. However, most crowdsourced social media data is unstructured and noisy, requiring continuous filtering and the establishment of minimum acceptable utility thresholds. AI can also be used to automatically categorise, geolocate and translates reports from formal county-level citizen engagement efforts, like public participation forums and call centres, to analyse and address public feedback. Government decision-makers can use these insights for crisis response, tracking service delivery and identifying emerging community needs. Counties are already exploring this. For example, Kilifi plans to integrate social media insights in official reporting tools.

Data sources: Unstructured crowdsourced data collected directly from citizens (text and voice); social media and digital listening data; county-level public participation and feedback records.

Feasibility: Extracting personal views from public platforms carries ethical and privacy risks, requiring strict data minimisation processes. Coverage also suffers from a digital participation bias, leaving low-connectivity populations underrepresented. While social media platforms offer structured API access, most formal government participation records are stored as unstructured PDF documents or physical meeting minutes that are not machine-readable.

GSMA Head Office
1 Angel Lane
London
EC4R 3AB
United Kingdom

gsma.com

GSMA Mobile for Development

M4D