The Map Is Not the Terrain:

Digital Public Infrastructure and the Data Foundations for AI-Enabled Government

Bryson Mwamburi, Genesis-Analytics. (2026)
The Map Is Not the Terrain
Digital Public Infrastructure and the Data Foundations for AI-Enabled Government

G:ENESIS
UNLOCKING VALUE

Digital Public Infrastructure and the Data Foundations for AI-Enabled Government

The foundations for public-sector AI must begin beneath the model layer. Before asking what an AI system can infer, recommend or automate, countries need an information environment capable of providing a sufficiently reliable representation of the society in which that system will operate.

Context: AI is changing what makes national data valuable

Every government holds a digital representation of the country it governs. Civil registries represent people and vital events; land systems represent property; tax systems represent economic activity; health systems represent patients and facilities; education systems represent learners and schools. Together, these administrative systems form an imperfect map of the people, institutions, places and relationships through which the state understands and serves society.

That map is thinnest exactly where the world is youngest and growing fastest. The World Bank estimates that nearly half of Africa's population lacks a legal identity, a direct consequence of civil registration and vital statistics systems that remain, region-wide, the weakest of any part of the world.[1] South Asia tells a related story from a different starting point: birth registration there has climbed faster than in any other region over the past two decades, from 39% in 2008 to 76% in 2024, and more than 42 million children under five in the region are still unregistered today.[2] Different regions, different trajectories, the same underlying exposure - a person who was never registered at birth does not later become visible to an AI system reasoning over administrative data; they simply never appear on the map at all.

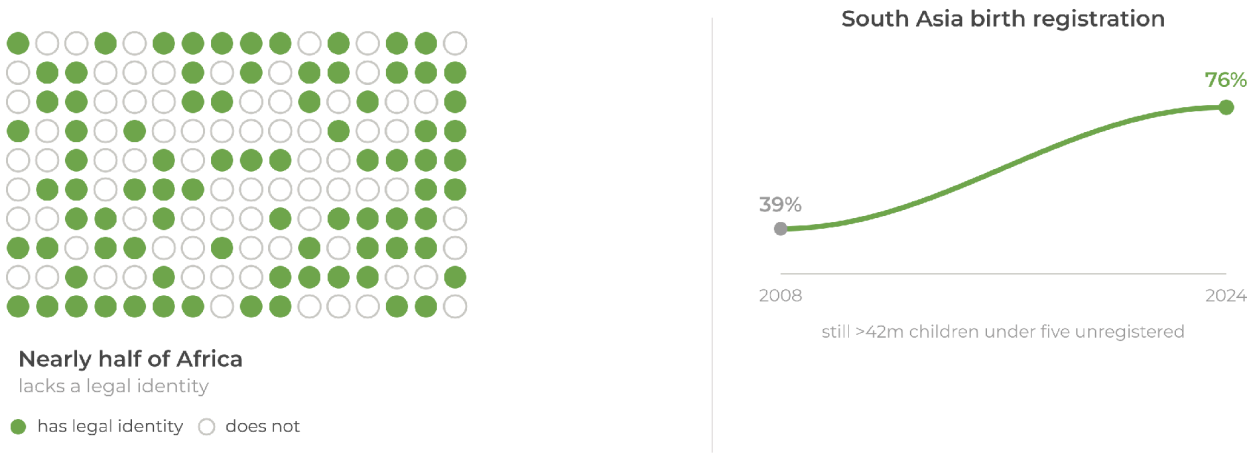


Figure 1. The identity gap across two young regions: roughly half of Africa lacks legal identity; South Asia's birth-registration rate has climbed from 39% to 76% since 2008 but still leaves over 42 million children uncounted.

Artificial intelligence makes the quality of this map considerably more consequential. Governments can increasingly access sophisticated reasoning, language and analytical capabilities without developing frontier models themselves. Commercial models, open-source ecosystems and shared platforms are lowering the barriers to capable machine intelligence. The models will continue to change, but access to intelligence is likely to become progressively less distinctive.

What remains stubbornly local is the context that allows this intelligence to understand a country. A general-purpose model does not arrive at knowing which businesses are currently registered, where public services exist, which households receive which benefits, whether a particular land record is authoritative or how a citizen appears across different government systems. That understanding must come from the country's own information environment.

[1] *World Bank / UNECA, "Data and Statistics: What Are the Current Gaps within DITE Africa?" (2025).*

[2] *UNICEF South Asia, The Right Start in Life: Global Levels and Trends in Birth Registration, 2024 update*

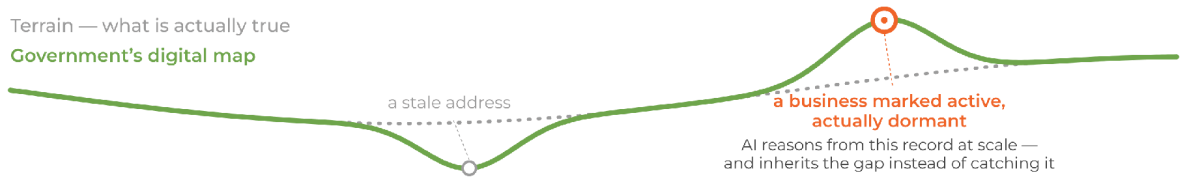
Models may increasingly be imported; the map they need to understand a country cannot be. This distinction should influence how countries think about the foundations for AI. Compute, data centres, connectivity, electricity and the water and cooling required to sustain them plainly matter and will determine what countries can host and operate.
Immediately above this physical infrastructure, however, lies another foundation: the ability of the state to produce information that AI can reliably interpret and use.

The strategic value of national data may therefore increase as intelligence itself becomes more widely available. The advantage will lie less in simply possessing large quantities of data and more in establishing which information can be trusted, connecting it appropriately across institutions and preserving its meaning as it moves.

The problem: the map is not the terrain

AI does not encounter a country directly; it encounters the representation of that country made available through data. The distinction, captured in Alfred Korzybski's observation that "the map is not the territory", is particularly important for public-sector AI. Administrative data describe reality, but they are not reality itself. They reflect when, why and how institutions collected information, alongside the limitations of the systems and processes through which it was recorded. This piece uses terrain rather than Korzybski's territory from here on deliberately: territory is the abstraction a map claims to represent,while terrain is what a system actually has to cross on the ground, one record at a time — which is why its quality matters so much.

Government's digital map is inevitably imperfect. A citizen may have an updated address in one database and an old address in another. A business can appear active according to one government system and dormant according to another. A health facility may have changed classification without that change reaching the systems used for planning or financing. Some populations and activities are richly represented in administrative systems, while others remain partially visible or absent altogether.



.

Figure 2. Where the map drifts from the terrain: a government's digital record mostly tracks reality but drifts at points of change, and an AI system reasoning at scale inherits that drift rather than catching it.

Human bureaucracies have traditionally compensated for these imperfections through judgement and institutional knowledge. Experienced officials recognise anomalies, understand which records tend to be more current and know whom to telephone when two systems disagree. Much of this context sits around the data rather than within it.

South Africa's SRD grant shows what this looks like once an automated system is already load-bearing. The grant - R370 a month, introduced during the pandemic for unemployed adults with no other income - added an automated bank-verification check in 2022: any inflow above a low threshold, regardless of source or timing, triggers ineligibility.

A 2025 survey by the Institute for Economic Justice found that only 10.3% of eligible respondents were actually receiving the grant - an erroneous exclusion rate of 89.7% - and traced 80% of rejections to that single automated check, which the researchers estimate should properly have applied to roughly a quarter of cases.[3] Nothing about the underlying eligibility rules changed. What changed was that an automated system now applied a blunt proxy at a scale and a speed no caseworker ever operated at, and the burden of proof shifted from the state to the claimant by default.

AI changes the consequences of an imperfect map because it can reason from it at an extraordinary scale. A human official may recognise that a record appears implausible before making a decision. An automated system processing hundreds of thousands of cases may reproduce the same erroneous assumption repeatedly. Greater intelligence does not correct an inaccurate representation of reality by itself; it can make the consequences of that inaccuracy larger. A foundation problem that starts out as a legal or governance gap does not stay contained there - the same failure mode in a data-integrity or interoperability layer, with an AI system reasoning across it at scale, is the harder version of the same story.

The foundations for public-sector AI must therefore begin beneath the model layer. Before asking what an AI system can infer, recommend or automate, countries need an information environment capable of providing a sufficiently reliable representation of the society in which that system will operate.

[3]Institute for Economic Justice survey, cited in “How is South Africa's welfare algorithm failing the poor?”, Context / Thomson Reuters Foundation, 30 June 2025.

Framing the foundation: integrity and interoperability

Two capabilities determine whether the state's digital representation can provide useful context for AI: data integrity and data interoperability. Integrity determines the fidelity of the map. Interoperability determines whether its separate sheets can be joined without losing their meaning.

Data integrity establishes whether information provides a defensible representation of reality. Governments already possess unusually valuable information through civil registration, taxation, health, education, land administration, licensing and social protection. Yet possessing a record does not make it authoritative, current or correct.

This is a statement about where the most authoritative records currently sit, not an argument for centralising data in the state rather than opening it. The integrity and interoperability problems described here apply equally to a government's own registries and to the open datasets it publishes - an open dataset built on an unreliable underlying record is no more trustworthy for having been made public.

Integrity therefore extends beyond conventional measures of data quality. Accuracy and completeness matter, but so do provenance, timeliness, lineage and institutional authority. An AI system needs enough context to distinguish an authoritative record from a convenient one, and a current record from one that was once accurate but has become obsolete.

This makes basic data-management capabilities foundational to reliable AI. Authoritative registries, common identifiers, metadata, stewardship, audit trails and mechanisms for correcting records may appear less sophisticated than the intelligence layer above them. They establish the information environment against which machine inference can be tested. The African Development Bank estimates the economic potential of open,well-governed data for Africa at roughly 1–2% of the region's GDP - value that stays uncaptured for exactly the reason this piece describes: the underlying registries are not yet reliable or connected enough to realise it.[4]

Chile's experience with its child-welfare risk-scoring system, Sistema Alerta Niñez, shows the same underlying problem in a different shape: an independent audit found the model was effectively weighting poverty indicators rather than safety indicators, so it flagged poor neighbourhoods as high-risk rather than unsafe ones - directing intervention toward the wrong households at national scale.[5] The record wasn't missing or unreconciled, as in Kenya or South Africa. It was simply measuring the wrong thing with complete confidence.

Interoperability determines how far trusted information can extend beyond the institution that created it. Most public problems cross institutional boundaries. A farmer can simultaneously exist within identity, land, agriculture, taxation, financial and
social-protection systems. A child can appear across civil registration, healthcare, education and welfare. Each institution may hold an accurate piece of the picture without any one institution holding enough context to understand the whole.

[4]Mo Ibrahim Foundation / African Development Bank, The Power of Data for Governance: Closing Data Gaps to Accelerate Africa's Transformation (2023 Ibrahim Index of African Governance report).
[5]Independent audit findings on Sistema Alerta Niñez, cited in Center for Human Rights and Global Justice, "Risk Scoring Children in Chile" (2022); catalogued in "The Fall of an Algorithm: Characterizing the Dynamics Toward Abandonment" (2024).

An AI system confined to one institutional dataset inherits that institution's partial view of reality. Moving from isolated AI applications towards genuinely joined-up public services therefore requires mechanisms through which relevant information can be assembled appropriately across government.

Useful interoperability must preserve meaning rather than merely move records. Two databases can exchange information successfully while interpreting it differently. A household, beneficiary, business or geographical unit may have different operational meanings across institutions. Apparently identical fields may have been collected for different purposes or refreshed at different intervals.

Semantic interoperability becomes more important when machines assemble context themselves. Human officials routinely infer meaning from institutional conventions and recognise when apparently similar information should not be treated as equivalent. AI requires more of this context to be explicit in standards, schemas, metadata and shared definitions.

Integrity and interoperability together create what we describe as a trusted national context. Integrity without interoperability produces islands of reliable information.
Interoperability without integrity can distribute unreliable information more efficiently. Their combination allows authoritative information to travel while retaining enough meaning, provenance and governance to be interpreted responsibly.

At Genesis, we've found it useful to track these questions through the GOAT framework - governance, operations, adoption and technology - precisely because integrity and interoperability failures rarely show up as a single clean technical bug. They show up as a governance gap, an operational workaround, or a technology choice made years before anyone thought to ask what AI would eventually need from it.

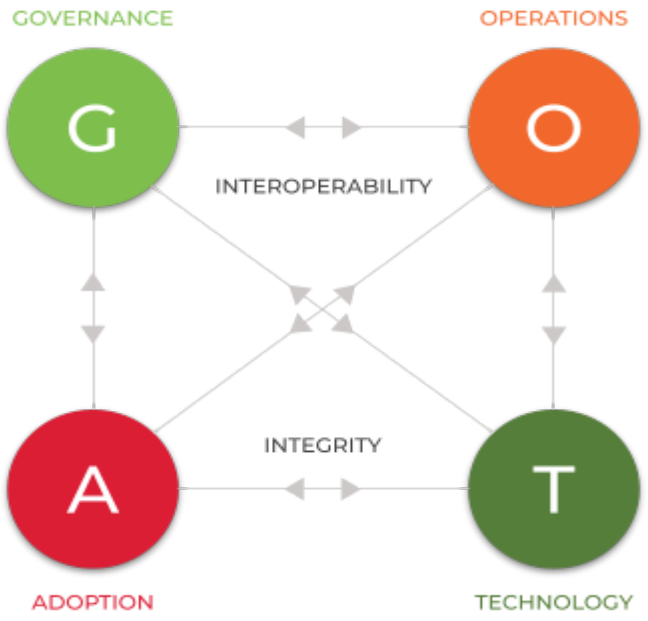


Figure 3. The GOAT framework: governance, operations, adoption and technology, with integrity and interoperability cutting across all four dimensions.

DPI: connecting the map

Digital Public Infrastructure provides much of the architecture through which trusted national context can become a reusable national capability. The DPI community has spent years developing common building blocks for digital systems: identifiers and registries, data exchange, trust infrastructure, common standards, payments and other shared capabilities. These foundations become increasingly valuable as AI begins to consume information across them.

DPI can be understood, in part, as providing the cartography of the digital state. Registries establish authoritative reference points. Identifiers help establish that records held in different places describe the same person, business or entity. Standards create common ways of representing information. Data exchange allows relevant information held across institutions to be connected without requiring every dataset to be assembled centrally.

India's identity and payments stack is the clearest illustration of what this looks like at genuine national scale. Aadhaar, the country's biometric identity system, had recorded over 150 billion cumulative authentication transactions by April 2025, with more than 2 billion further authentications happening most months. Built on top of that identity layer, the UPI payments rail processed over 20 billion transactions in a single month in August 2025, and has since expanded into seven other countries.[6][7] That is what interoperable identity and payments infrastructure can do once it is actually connected.

It is also, in the same breath, the clearest illustration of why integrity matters as much as scale. An IDinsight study for the Omidyar Network found that biometric authentication failures had denied more than two million people access to their legally entitled food rations across just three states.[8] India's Right to Food Campaign has documented dozens of hunger-related deaths since 2015 in which Aadhaar authentication failure was a contributing factor.[9] Neither finding is an argument against interoperable national infrastructure - India's stack has done more to connect its map than almost any system in the world. It is an argument for exactly the point this piece keeps returning to: interoperability without integrity does not fail safely. It fails at the scale the infrastructure was built for.

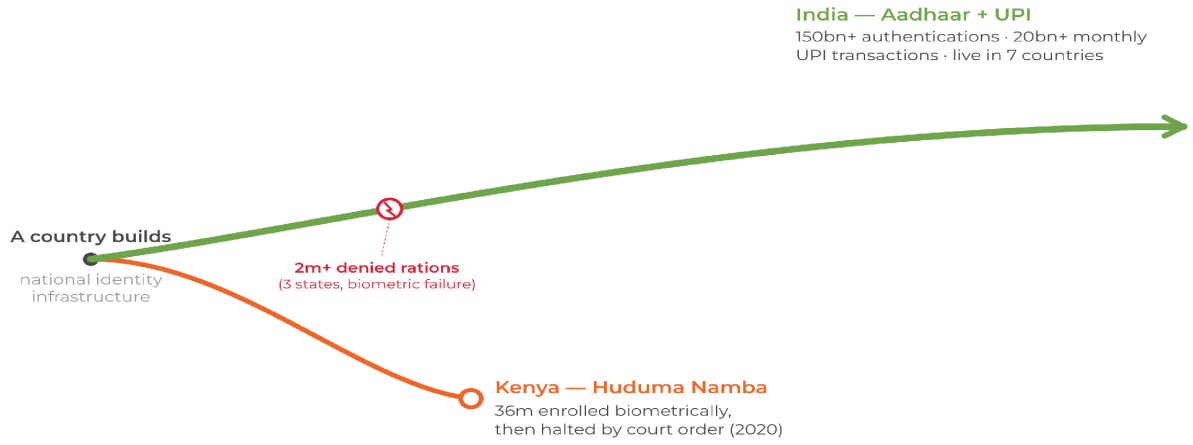


Figure 1. Two national identity systems, two outcomes: Kenya's rollout was halted by court order in 2020; India's scaled to over 150 billion authentications but still shows a documented integrity gap.

[6] *UIDAI, cumulative Aadhaar authentication statistics, April 2025.*
[7] *National Payments Corporation of India (NPCI), UPI transaction volumes, August 2025.*
[8] *IDinsight / Omidyar Network, State of Aadhaar 2017–18.*
[9] *India's Right to Food Campaign, documentation of Aadhaar-linked exclusion deaths.*

This interpretation builds from rather than replaces established DPI thinking. CDPI's work places identifiers and registries, data sharing, trust infrastructure and interoperable standards among the foundations of digital ecosystems, while its emerging work on AI-ready DPI recognises that these capabilities will increasingly interact with machine intelligence. The OECD similarly identifies high-quality interoperable data and coherent DPI as important foundations for scaling trustworthy AI in government.

AI gives these established capabilities another strategic purpose. Infrastructure designed principally to make digital services interoperable can also become infrastructure through which AI assembles the context needed to reason about people, institutions and services.

The data-exchange layer becomes particularly important as intelligence moves across institutional boundaries. Its role is no longer limited to transmitting records from one government service to another. It increasingly becomes the means through which AI can assemble a coherent representation from information held across the state.

Machine reasoning raises the standard expected of that exchange. Infrastructure designed to transfer a known field between two systems was not necessarily designed for AI dynamically assembling context from several sources. Information must increasingly carry sufficient context to establish what it represents, where it originated, how current it is and the conditions under which it may legitimately be used.

The next frontier for data exchange is therefore preserving trustworthy context as information moves. This extends the DPI community's established emphasis on authoritative registries, federation, standards and semantic interoperability into an environment where exchanged information increasingly becomes material for machine reasoning.

The goal: a state that intelligence can understand

The ambition should not be to construct a perfect digital representation of society. No administrative information system can capture the full complexity of the people and communities it describes. The goal is instead to create an information environment sufficiently trustworthy, interoperable and governed for increasingly capable machines to reason from it responsibly.

This requires thinking beyond whether today's DPI is compatible with today's AI models. Models, vendors and architectures will change rapidly. High-integrity administrative data, authoritative registries, shared semantics and governed interoperability are more durable assets because they remain useful as the intelligence operating above them evolves.

The emerging public-sector information chain can be understood as a progression from data to service:

Data → Trusted Context → Intelligence → Decision → Service

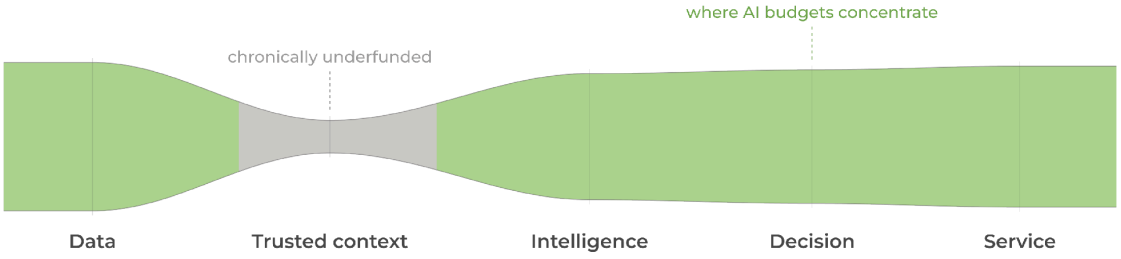


Figure 2. The public-sector AI pipeline: investment concentrates in intelligence, decision and service, while the trusted-context stage in between remains chronically underfunded.

Each layer depends on the foundations beneath it. Administrative systems and registries provide data. Integrity and interoperability turn those data into trusted context. AI interprets that context and supports decisions. Identity, credentials, payments, notifications and service platforms allow those decisions to become legitimate public services.

The citizen experience shows why the sequence matters. Consider someone who has recently lost her livelihood and approaches the government for assistance. A conventional digital service can help her establish her identity, locate a relevant programme and submit an application. A more intelligent service could recognise her circumstances, identify programmes for which she may qualify, reuse information the government legitimately holds, identify missing evidence and route her towards an appropriate intervention.

The intelligence layer alone cannot deliver that experience. The service must resolve the citizen correctly, retrieve authoritative information, connect records held by different institutions, preserve their meaning and provenance and determine which information can legitimately inform a decision. Without those foundations, AI risks providing a sophisticated interface over the same fragmented bureaucracy.

Investment in trusted national context is therefore unusually durable in a fast-moving technology landscape. Strengthening registries, resolving identifiers, improving administrative data, agreeing shared semantics and developing governed data exchange improve public services today while increasing what successive generations of AI can reliably do tomorrow.

From the map to implementation

Everything above returns to one claim: the foundations have to be built before the model layer, not alongside it or after it. Building trusted national context is not solely a data architecture challenge, either. Better standards and exchange mechanisms matter, but the quality of the information environment also depends on how institutions govern data, how public services operate, whether people trust new systems and how technology is implemented.

GOAT is useful here because it prevents AI readiness from becoming another

technology-readiness exercise. A sophisticated exchange architecture achieves little when institutions cannot agree which source is authoritative or have no operational mechanism for correcting inaccurate records. Equally, sound governance and enthusiastic adoption cannot create joined-up services where systems remain unable to understand one another.

Kenya's experience with the Huduma Namba is what the governance dimension looks like when it's missing, and it didn't even require an AI system to go wrong. Between April and May 2019, the government collected biometric data from an estimated 36 million people to build the National Integrated Identity Management System, before a data protection law existed to govern how that data would be held, shared or corrected. Kenya's High Court halted the rollout in January 2020, finding the legal framework "inadequate and totally wanting," [10] and specifically flagged the risk that communities without existing documentation - among them the Nubian and Somali minority populations - would be excluded from a system built to be mandatory for accessing government services.

No AI system was reasoning over that data yet. Had one been added on schedule, it would have inherited precisely the governance gap the court identified, at whatever scale the system was deployed. That is the practical argument for treating governance as an AI-readiness question from the outset, rather than a legal track that can be resolved in parallel while the technology ships on time.

The most important questions now sit between these dimensions. Governments will need practical approaches for resolving conflicts between authoritative sources, establishing how much semantic interoperability machine reasoning requires, maintaining meaningful human oversight as AI draws on more data sources and governing access when the consumer of public information is increasingly a machine.

The investment choices will also become harder as AI capabilities advance. Governments will need to judge when another improvement in model capability produces less public value than improving the quality, accessibility or interoperability of the information beneath it. For many public-sector applications, better context may ultimately prove more valuable than better intelligence.

[10] *Kenya High Court judgment in Nubian Rights Forum & Katiba Institute v. Attorney General (the NIIMS case), 30 January 2020.*

Partnership: improving the map together

The next stage of AI-enabled government will be shaped through implementation rather than technology discourse alone. The principles need to survive contact with real registries, institutional boundaries, legacy systems, public servants and the imperfect administrative information that governments actually possess.

No single community is positioned to answer these questions independently. AI practitioners understand rapidly evolving model capabilities. The DPI community brings deep experience in reusable infrastructure, standards and interoperable ecosystems. Governments understand the institutional and political realities within which public services operate. Development partners bring experience across countries and systems.

The opportunity lies in bringing these perspectives together around real public problems. This means identifying where weak integrity distorts the state's representation of reality, testing where interoperability genuinely improves service delivery, understanding what context machines require and developing practical approaches to governance as AI becomes more deeply connected to operational government systems.

This is where the GOAT lens introduced earlier stops being a framework and starts being a mandate. For Genesis, closing these gaps is a natural extension of our work across data, digital transformation and public-sector problem solving; GOAT gives us one way of connecting the governance, operational, adoption and technology dimensions of the challenge. Our experience designing data and AI solutions gives us another: the ability to follow the argument from the quality of an individual record through to the decision or service that ultimately depends upon it.

For me, working at the intersection of data science, AI solutions and public-sector transformation, the opportunity is to move beyond demonstrating what AI can do towards understanding what must be true around it for those capabilities to create lasting public value. The difficult work is increasingly in the connections: between datasets, between institutions, between technology and operations, and ultimately between an AI-generated inference and the reality it purports to describe.

Genesis and I want to be partners in answering these questions as governments and the wider DPI community navigate this emerging space. That means testing the propositions through implementation, learning where existing architectures succeed or fail and helping define what trusted national context needs to look like as machines play a larger role in public-service delivery.

The map will never become the terrain, nor should we expect it to

G:ENESIS
UNLOCKING VALUE