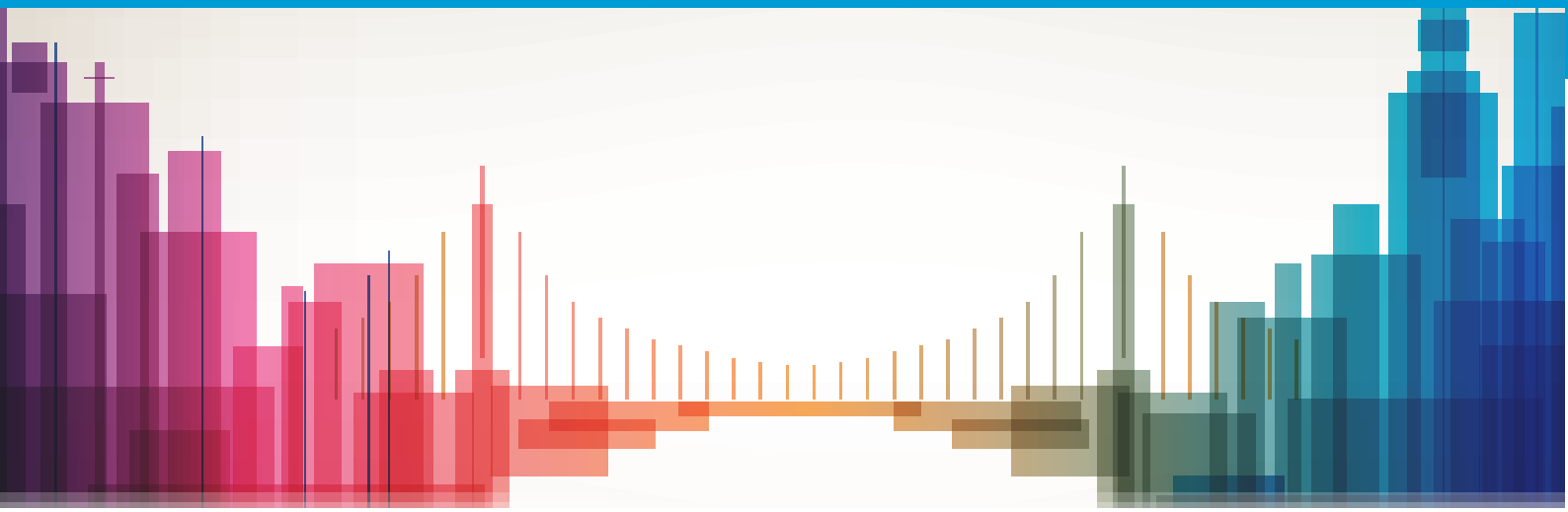
AI-Driven urban key performance indicators (KPIs)

Harnessing non-traditional data and artificial intelligence to measure city performance with unprecedented accuracy, granularity and timeliness



Convention on Biological Diversity

United Nations ECLAC

Food and Agriculture Organization of the United Nations

ITU

UN-HABITAT

UNDP

United Nations Economic Commission for Africa

UNECE

UNIDO

unesco

UN environment programme

FINANCE UNEP INITIATIVE

United Nations Framework Convention on Climate Change

unitar United Nations Institute for Training and Research

UNFPA

United Nations | Department of Economic and Social Affairs

UNJSPF United Nations Joint Staff Pension Fund

UN OFFICE 4 PARTNERSHIPS

UNU EGOV

UN Tourism

WORLD METEOROLOGICAL ORGANIZATION


AI
CO_2

AI-Driven urban key performance indicators (KPIs)

Harnessing non-traditional data and artificial intelligence to measure city performance with unprecedented accuracy, granularity and timeliness

Foreword

This publication was developed within the framework of the United for Smart Sustainable Cities (U4SSC) initiative.

Acknowledgements

The development of this deliverable was led and coordinated by Okan Geray (Dubai Digital Authority, United Arab Emirates).

This publication was authored by Okan Geray (Dubai Digital Authority, United Arab Emirates).

The author would like to thank the following contributing individuals: Cristiano Passerini (Universita di Bologna), David Pastor (Universidad Politecnica de Madrid), Jennifer Perez (PolicySimAI), Najeeba Al Marzooqi (Dubai Data and Statistics Establishment), Payal Agarwal (PolicySimAI), Shabnam Sheikh (Strategic Researcher), Silvia Blanco (Freelance AI Expert), Swati Dhiman (University College London), and Yiwei Song (Independent Expert).

The author wishes to thank the U4SSC management team: Ramy Ahmed Fathy, Giampiero Bambagioni, Paolo Gemma, Wendy Teresa Goico Campagna, Tania Marcos and Emily Royall (U4SSC Vice-Chairs) for their assistance and contributions.

The author also extends gratitude to the contributing organizations along with their representatives: Oliver Hillel from the Convention on Biological Diversity (CBD), Lucy Winchester and Vera Kiss from the Economic Commission for Latin America and the Caribbean (ECLAC), Nevena Alexandrova from the Food and Agriculture Organization (FAO), Cristina Bueti from the International Telecommunication Union (ITU), Deniz Susar from United Nations Department of Economic and Social Affairs (UNDESA), Iryna Usava from the United Nations Development Programme (UNDP), James Murombedzi from the United Nations Economic Commission for Africa (UNECA), Dario Liguti and Tea Aulavuo from the United Nations Economic Commission for Europe (UNECE), Guilherme Canela from the Regional Bureau for Sciences in Latin America and the Caribbean of the United Nations Educational, Scientific and Cultural Organization (UNESCO), Gulnara Roll and Mrityunjay Sharma from United Nations Environment Programme (UNEP), Matthew Ulterino from the United Nations Environment Programme Finance Initiative (UNEP-FI), Motsomi Maletjane from the United Nations Framework Convention for Climate Change (UNFCCC), Soren Thomassen from United Nations Population Fund (UNFPA), Aline Machado da Matta and Isabel Wetzel from the United Nations Human Settlements Programme (UN-Habitat), Katarina Barunica Spoljaric and Nicholas Dehod from the United Nations Industrial Development Organization (UNIDO), Ebru Canan-Sokullu from the United Nations Institute for Training and Research (UNITAR), Dino Cataldo Dell'Accio from United Nations Joint Staff Pension Fund (UNJSPF), William Kennedy from the United Nations Office for Partnerships (UNOP), Soumaya Ben Dhaou from the United Nations University – Operating Unit on Policy-Driven Electronic Governance (UNU-EGOV), Roberta Boscolo from World Meteorological Organization (WMO) and Sandra Carvao from the World Tourism Organization (UN Tourism).

Disclaimer

The opinions expressed in this publication are those of the authors and do not necessarily represent the views of their respective organizations or U4SSC members. In line with the U4SSC principles, this report does not promote the adoption and use of Smart City technology. It advocates for policies encouraging responsible use of information and communications technologies (ICTs) that contribute to the economic, environmental and social sustainability as well as the advancement of the 2030 Agenda for Sustainable Development and the Pact for the Future and its Global Digital Compact.

ISBN

978-92-61-43741-1 (electronic version)
978-92-61-43751-0 (EPUB version)



This work is licensed to the public through a Creative Commons Attribution-Non-Commercial-Share Alike 3.0 IGO license (CC BY-NC-SA 3.0 IGO).

For more information, please visit https://creativecommons.org/licenses/by-nc-sa/3.0/igo/

© CBD, ECLAC, FAO, ITU, UNDESA, UNDP, UNECA, UNECE, UNESCO, UNEP, UNEP-FI, UNFCCC, UNFPA, UN-Habitat, UNIDO, UNITAR, UNJSPF, UNOP, UNU-EGOV, WMO and UN Tourism.


25.81
97.84
59.85

Contents

Abbreviations and acronyms viii

Executive Summary ix

1 Introduction: Why Cities Must Reinvent Measurement 1
- 1.1 The measurement gap in modern city management 1
- 1.2 A new possibility: The continuously measured city 1
- 1.3 Purpose and structure of this document 2

2 The Non-Traditional Data Landscape 3
- 2.1 Satellite and aerial earth observation 3
- 2.2 Drone and UAV imagery 4
- 2.3 Mobile network and location data 4
- 2.4 IoT sensor networks 4
- 2.5 Connected vehicles and mobility platforms 4
- 2.6 Street-level imagery and video 5
- 2.7 Transactional and administrative digital exhaust 5
- 2.8 Social media, web and crowdsourced data 5

3 The AI Toolkit: From Raw Signals to Trusted Indicators 6
- 3.1 Computer vision and deep learning 6
- 3.2 Supervised machine learning for estimation and inference 7
- 3.3 Time-series forecasting and predictive analytics 7
- 3.4 Natural language processing (NLP) and large language models (LLMs) 7
- 3.5 Data fusion and multimodal learning 7
- 3.6 Edge AI and real-time processing 8
- 3.7 Digital twins and simulation 8
- 3.8 Generative and agentic AI 8

4 From Data to Decision: The AI-Driven KPI Value Chain 9

5 Where It Works: Application Domains Across the City 11
- 5.1 Mobility and transportation 12
- 5.2 Environment, climate and energy 12
- 5.3 Urban planning, land use and housing 12
- 5.4 Infrastructure and utilities 13
- 5.5 Public safety, resilience and emergency management 13
- 5.6 Economy, tourism and urban vibrancy 13
- 5.7 Health, wellbeing and liveability 13
- 5.8 Governance and service delivery 14

6 From Concept to Practice: Case Studies 15
- 6.1 Population: Dubai Population Now (United Arab Emirates) 16

6.2 Health: dengue forecasting in Ho Chi Minh City and Ha Noi (Viet Nam) 16
6.3 Mobility and accessibility: Project Sidewalk 17
6.4 What the cases show 18

7 The Value Proposition: Why This Matters 19

8 Navigating the Challenges: Risks and Safeguards 20
8.1 Privacy and social licence 21
8.2 Bias and representativeness 21
8.3 Accuracy, validation and model drift 22
8.4 Interoperability, legacy systems and vendor lock-in 22
8.5 Skills, institutions and regulation 22

9 A Roadmap for Aspiring Cities 22
9.1 A maturity model for AI-driven measurement 23
9.2 A phased implementation path 24

10 Conclusion: From Periodic Reports to a Self-Aware City 24

List of figures, tables and box

Figures

Figure 1: Form Data to Impact: AI-Driven Urban KPIs xi
Figure 2: From periodic reporting to continuous, granular and predictive urban measurement xii
Figure 3: Eight non-traditional data families that can feed urban KPI production 3
Figure 4: The AI-driven KPI value chain and its three permanent safeguards 10
Figure 5: Application domains for AI-driven urban KPIs 11
Figure 6: Core risk areas and corresponding safeguards for trusted AI-driven KPIs 21
Figure 7: Phased implementation path for aspiring cities 23

Tables

Table 1: The non-traditional data landscape at a glance 5
Table 2: The AI toolkit mapped to its role in KPI production 8
Table 3: What changes when KPIs move from periodic collection to continuous computation 10
Table 4: Consolidated sample KPI catalogue by domain – a menu from which aspiring cities can select their first lighthouse indicators 14
Table 5: The three case studies at a glance 15
Table 6: Five levels of measurement maturity. Levels 1–3 are achievable by any committed city within two years; levels 4–5 describe the current global frontier 23

Box

Box 1: Before measuring: checking data readiness (Bologna, Italy) 19

Abbreviations and acronyms

Abbreviation	Full form
AI	Artificial intelligence
API	Application programming interface
AQI	Air Quality Index
CO2	Carbon dioxide
CRM	Customer relationship management
CV	Computer vision
EO	Earth observation
EU	European Union
GNN	Graph neural network
IoT	Internet of Things
ISO	International Organization for Standardization
ITU	International Telecommunication Union
KPI	Key performance indicator
LLM	Large language model
ML	Machine learning
MLOps	Machine learning operations
NLP	Natural language processing
PM2.5	Fine particulate matter with a diameter of 2.5 micrometres or less
POI	Point of interest
RNN	Recurrent neural network
SAR	Synthetic aperture radar
SDG	Sustainable Development Goal
U4SSC	United for Smart Sustainable Cities
UAV	Unmanned aerial vehicle
UN	United Nations

Executive Summary

Cities have always measured performance. The difference now is that they can be measured continuously, comprehensively and predictively without waiting for the next census, survey or manual count.

Most cities are run on old numbers. Population figures can be years out of date, traffic and air quality are measured at a handful of points, and results often arrive long after the decision they were meant to inform. This report shows how cities can combine artificial intelligence with data they largely already generate, from satellites, sensors, mobile networks and their own service records, to measure how they are performing faster, in finer detail and at a lower cost per measurement.

What this offers city leaders

- **Know sooner.** Indicators that used to arrive once a year can be refreshed daily or hourly, and some can be forecast before the event.
- **See every neighbourhood.** Results at street and block level show which areas have worse air, slower services or less accessible streets. City-wide averages hide these differences.
- **Act before problems grow.** Forecasts of disease outbreaks, pollution episodes or infrastructure failures allow a city to prepare instead of react.
- **Measure more for the same budget.** Once the systems are in place, each additional measurement costs far less than a new survey or manual count. The systems, data licences and skilled staff still need sustained funding.

What the evidence shows

The report presents three implementations. In Dubai, linking government registers raised the official population figure for the end of 2024 by about 10 per cent, or roughly 385 000 people, and the figure is now updated continuously. In Viet Nam, a dengue forecasting system issued warnings up to six months ahead and reduced forecast error by about 18 per cent against a seasonal baseline over 39 months of live operation. Across 35 cities, volunteers supported by AI have mapped sidewalk barriers street by street; their results agreed with government field audits between 75 and 100 per cent of the time on which features exist, but much less often on their condition.

Built on the indicators cities already report

AI-driven measurement does not replace established indicator frameworks. It is a new way of producing the numbers they require. The U4SSC Key Performance Indicators for Smart Sustainable Cities and the Global Urban Monitoring Framework of UN-Habitat, which the United Nations Statistical Commission endorsed in 2022, define what a city should measure and make cities comparable. This report addresses how to measure it better. Cities should keep these frameworks as their reference so that AI-derived figures can be reported alongside official statistics.

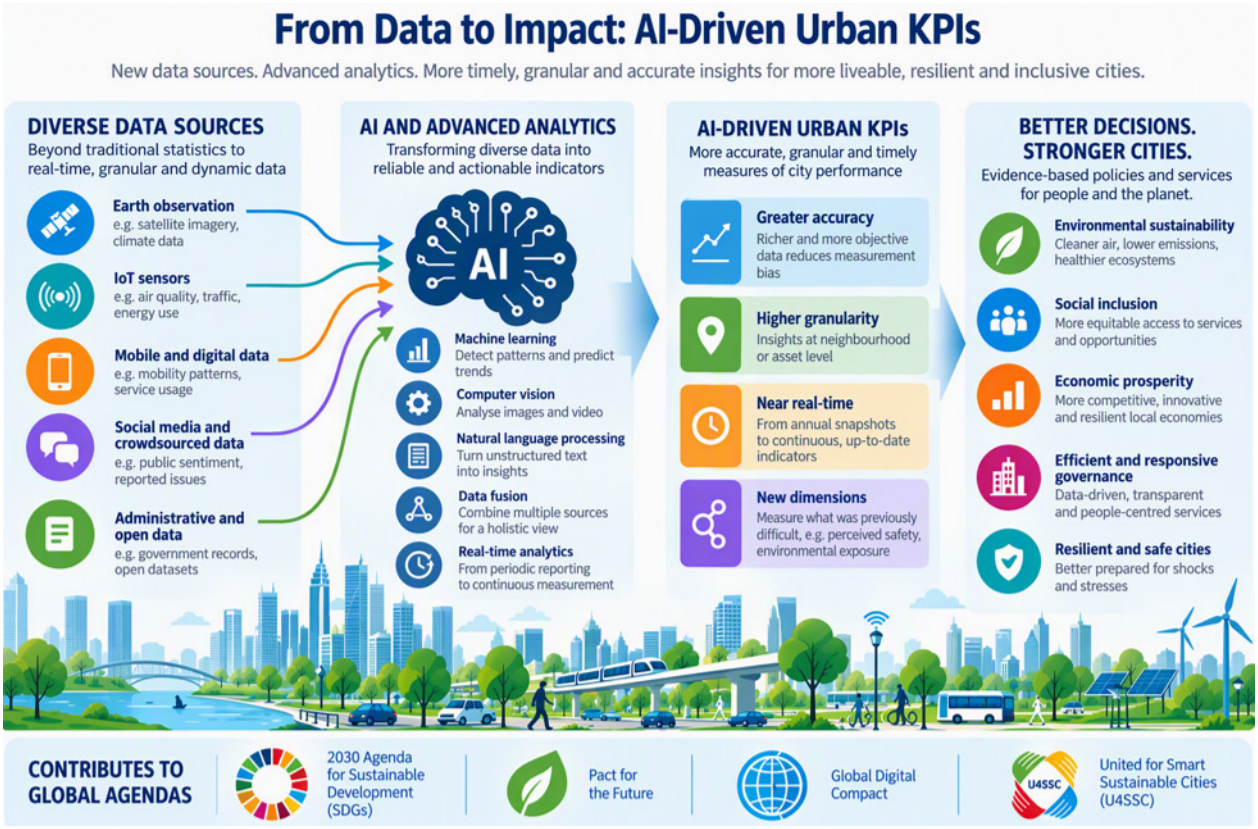
The risks, and how to manage them

These methods carry real risks: loss of privacy, bias against people who leave fewer digital traces, errors in sensors and models, and dependence on a single vendor. The report sets out the safeguards that go with each. AI-derived figures must be checked continuously against trusted ground measurements, their accuracy must be published, personal data must be aggregated and protected, and people must remain accountable for the decisions taken.

Where to start

A city does not need a large budget to begin. It can name one accountable lead, choose two or three indicators that a decision-maker will act on, start with free satellite imagery and data it already holds, run the new method alongside the old one, and publish the comparison. The report provides a five-level maturity model and a four-phase roadmap for cities of any size.

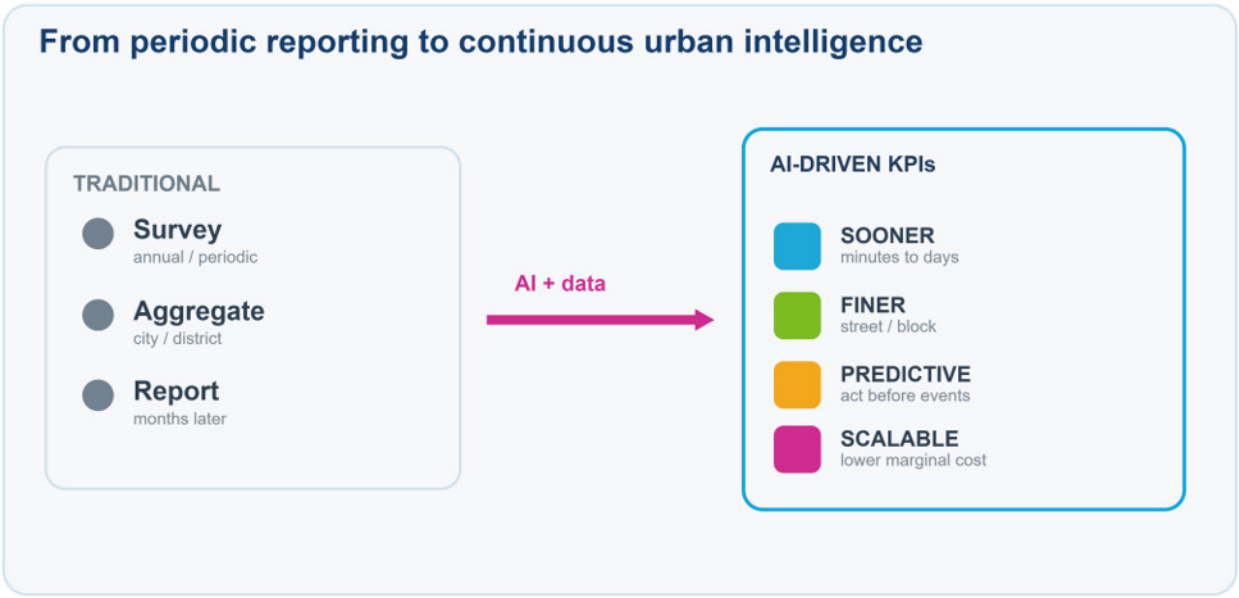
Figure 1: Form Data to Impact: AI-Driven Urban KPIs



Six questions for city leaders to ask

- **Decisions**. Which decisions would we take differently if we had this figure every week instead of every year?
- **Ownership**. Who is accountable for each indicator, and who will act on it?
- **Accuracy**. How do we know the figure is right, and do we publish how accurate it is?
- **Privacy**. Whose data are we using, how are they protected, and have residents been told?
- **Fairness**. Who is missing from the data, and how do we correct for that?
- **Independence**. Could we change supplier without losing our data and methods?

Figure 2: From periodic reporting to continuous, granular and predictive urban measurement




CHECKLIST DASHBOARD
Q1
Q2
Q3
Q4
JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC
DISCOVERY
PLANNING
MVP DEVELOPMENT
TIME BREAKDOWN
AI
TREND SCORE:
GROWTH RATE:
AVG. MONTHLY:
SALES DATA
CUSTOMER BEHAVIOR
MARKETING CAMPAIGN
WEBSITE TRAFFIC
REVENUE GROWTH
FINANCIAL DATA
100
90
80
70
60
50
40
30
20
10
0
JAN
FEB
MAR
APR
MAY
JUN
REVENUE
EXPENSES
NET PROFIT
60%
43%
85%
REVENUE GROWTH
MARKET PERFORMANCE
OPERATIONAL EFFICIENCY

1 Introduction: Why Cities Must Reinvent Measurement

1.1 The measurement gap in modern city management

Key performance indicators (KPIs) are how a city tracks its progress and is held accountable. Frameworks such as the Key Performance Indicators for Smart Sustainable Cities of the United for Smart Sustainable Cities initiative (U4SSC KPIs), the UN-Habitat Global Urban Monitoring Framework, and the UN Sustainable Development Goal 11 targets have given cities a shared vocabulary for what to measure, from commute times and air quality to green space per capita and emergency response performance. What these frameworks do not address is how to measure. In most cities, the underlying data still come from household surveys conducted every few years, traffic counts performed by hand or by temporary equipment, a sparse network of official monitoring stations, and administrative records that consolidate slowly across departments.

Two of these frameworks are the main reference points for this document. The U4SSC Key Performance Indicators for Smart Sustainable Cities, standardized in Recommendation ITU-T Y.4903, cover the economy, environment, and society and culture dimensions of a city. The Global Urban Monitoring Framework of UN-Habitat, endorsed by the United Nations Statistical Commission in March 2022, harmonizes existing urban indices into a single structure of five domains and four city objectives for monitoring the Sustainable Development Goals and the New Urban Agenda. AI-driven measurement is a way of producing the data for indicators in both frameworks. It does not introduce a competing set of indicators.

The result is a measurement gap with three dimensions:

- **Delay**. By the time an indicator is published, the condition it describes may have changed. A new highway interchange, a pandemic, a heatwave or a construction boom can reshape a city between two surveys.
- **Lack of local detail**. A city-wide average conceals the differences between neighbourhoods, which is where policy succeeds or fails. An acceptable average commute can hide districts in permanent gridlock.
- **Cost**. Because each data point is expensive to collect, cities ration measurement. They monitor a handful of locations and extrapolate to everywhere else.

1.2 A new possibility: The continuously measured city

Two developments have made a different model possible:

- **More data**. Cities today are covered by instruments that were never designed as statistical tools: earth-observation satellites revisiting every location on the planet every few days, billions of mobile devices generating anonymized movement signals, low-cost environmental sensors

on lamp posts and buses, cameras at intersections, smart meters in homes, fare cards on transit systems and the records left behind by online services.

- **Better tools to interpret it**. Modern machine learning can transform these raw, messy, high-volume signals into calibrated indicators that are reliable enough to base decisions on. It can count vehicles in video, outline buildings in satellite images, infer transport modes from movement traces, forecast pollution hours ahead and read the sentiment of thousands of service requests.

We define an AI-Driven Urban KPI in this document as a city performance indicator that is computed wholly or substantially by applying artificial intelligence techniques to non-traditional data sources, producing estimates that are more frequent, more spatially granular, more predictive or more cost-efficient than the traditional equivalent, while remaining validated against ground truth and governed by clear ethical safeguards.

In plain terms: an AI-driven urban KPI is a city indicator that software calculates from data the city is already producing. It is checked regularly against trusted measurements and is governed by clear rules that protect residents.

The emphasis on validation and governance is deliberate. An AI-driven KPI is not a replacement for statistical rigour; it is a new production method for statistics. The most successful cities treat traditional instruments such as surveys, reference-grade monitoring stations and official counts as the calibration backbone against which AI-derived estimates are continuously checked, in the same way national statistical offices have begun blending satellite and mobile data into official statistics.

1.3 Purpose and structure of this document

This is a concept and vision document for aspiring cities: those that want to understand what AI-driven measurement is, what it requires and what it makes possible, before committing to a programme. Section 2 catalogues the non-traditional data landscape. Section 3 explains the AI techniques that convert data into trusted indicators. Section 4 describes the end-to-end value chain and contrasts it with traditional measurement. Section 5 maps the application domains across city life, with illustrative KPIs for each. Section 6 profiles real city implementations, presented as three case studies together with a method for checking data readiness. Sections 7 and 8 weigh the benefits against the risks and safeguards. Section 9 offers a maturity model and phased roadmap, and Section 10 closes with a vision of the self-aware city.

How to read this report. City leaders with limited time can read the Executive Summary, Section 6 (case studies), Section 8 (risks and safeguards) and Section 9 (roadmap). Sections 2 to 4 give technical teams the detail on data sources and methods.

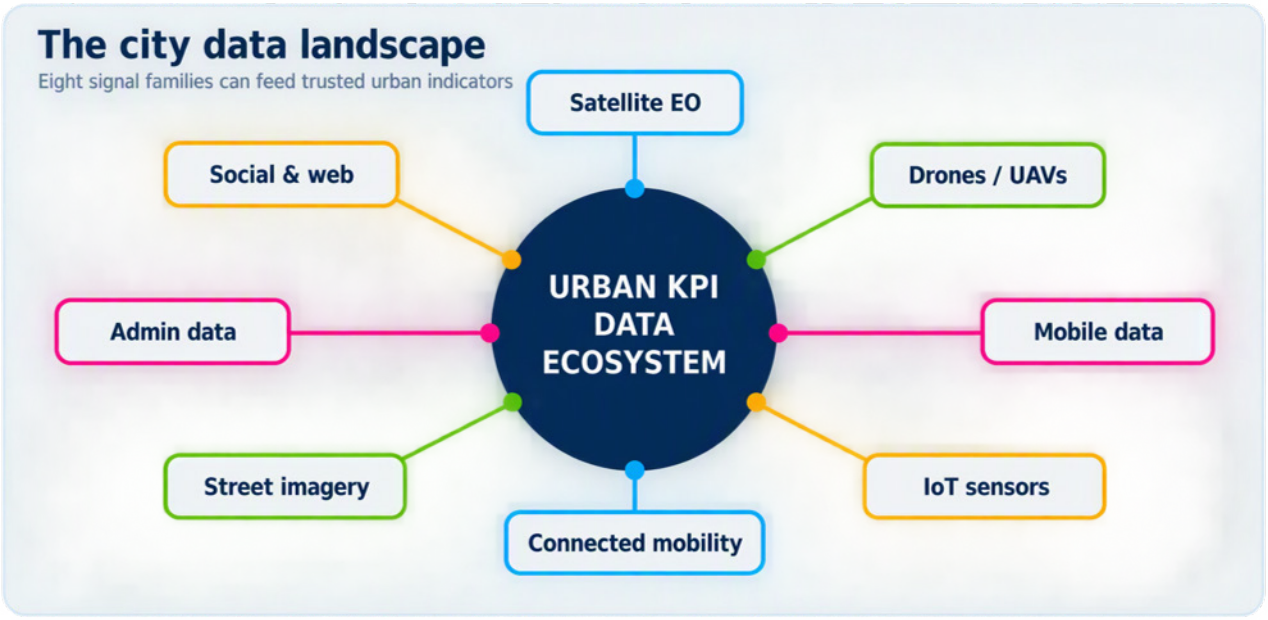
2 The Non-Traditional Data Landscape

The modern city emits a continuous stream of signals about itself. The craft lies in knowing what each source can and cannot say, and in combining them so their weaknesses cancel out.

In brief for city leaders. Your city already produces, or can obtain at low cost, much of the data described in this section. The practical questions are who holds each source, on what terms the city may use it, and what each source cannot tell you. Table 1 summarizes all eight sources on one page.

Non-traditional data, sometimes called alternative, unconventional or big data sources, refers to data generated as a by-product of other activities (communication, transport, commerce, earth observation) rather than collected purposely for statistics. Eight families of sources dominate urban applications today.

Figure 3: Eight non-traditional data families that can feed urban KPI production



2.1 Satellite and aerial earth observation

Earth-observation satellites are the only instrument that measures every city on the planet with identical, unbiased methodology. Open constellations such as the European Union's Sentinel

programme image most locations every five to six days at ten-to-twenty-metre resolution, free of charge; commercial constellations offer sub-metre imagery, some with daily revisit; and specialized sensors capture what the human eye cannot. Sentinel-5P maps nitrogen dioxide and other pollutants, thermal sensors reveal urban heat islands, radar (SAR or Synthetic Aperture Radar) sees through clouds and darkness to detect floods and ground subsidence, and night-time light sensors trace economic activity. For a city, this translates into indicators of built-up expansion, land-use change, green cover, construction progress, heat exposure, flood extent and informal settlement growth refreshed continuously rather than at each master-plan cycle.

2.2 Drone and UAV imagery

Where satellites provide breadth, drones provide depth. Unmanned aerial vehicles (UAVs) capture centimetre-scale imagery, thermal signatures and 3D point clouds on demand, making them the instrument of choice for asset-level indicators: facade and roof condition, vegetation health in parks and streetscapes, construction-site compliance, illegal dumping and post-storm damage.

2.3 Mobile network and location data

Every active mobile phone interacts continuously with the network, and when aggregated and anonymized under strict privacy controls, these interactions form the richest available picture of how people actually use a city. Telecom operator data can estimate dynamic population by grid cell and hour, origin-destination flows, commuting corridors, catchment areas of centres and events, tourism footfall and evacuation behaviour.

2.4 IoT sensor networks

The Internet of Things (IoT) gives the city a nervous system of purpose-built sensors: air-quality nodes on lampposts, noise meters, water-level gauges, smart electricity and water meters, waste-bin fill sensors, structural-health monitors on bridges and parking-occupancy detectors. Individually cheap and imperfect, in aggregate and after machine-learning calibration they achieve what a handful of reference stations never could: neighbourhood-scale, minute-by-minute coverage.

2.5 Connected vehicles and mobility platforms

Modern vehicles, transit fleets, ride-hailing platforms, bike-share and scooter systems continuously report speed, position and events. Floating-car data yields link-level travel times and congestion indices without a single roadside sensor; bus telemetry powers reliability and headway KPIs; micromobility data reveals where cycling infrastructure is missing. Transit smart-card records, millions of tap-ins and tap-outs, allow rigorous measurement of ridership, transfer patterns and service equity.

2.6 Street-level imagery and video

Cameras at intersections, on municipal vehicles and in street-view archives capture the city at eye level. Computer vision converts these feeds into counts and classifications of vehicles, pedestrians and cyclists; detections of incidents, potholes, damaged signs, broken streetlights and fallen trees; and systematic audits of walkability, shade and streetscape quality. Because analysis can run on the edge device and transmit only counts and events, modern designs can deliver these KPIs while discarding imagery and protecting privacy.

2.7 Transactional and administrative digital exhaust

Cities and their utilities generate torrents of operational records: municipal service requests (such as those made to the 311 non-emergency lines used in North America), permit and licensing transactions, point-of-sale aggregates, utility billing, smart-meter interval data and web or app telemetry from digital government services. These records, already collected for operations, can be repurposed, with appropriate anonymization, into indicators of service responsiveness, digital adoption, business dynamism and consumption efficiency at very little extra cost.

2.8 Social media, web and crowdsourced data

Public social-media posts, online reviews, property listings, job boards and citizen-reporting apps provide an unfiltered signal of how residents experience the city: sentiment about services, emerging complaints, prices and neighbourhood perception. Used carefully, with awareness of demographic skew, natural language processing turns this stream into early-warning and satisfaction indicators that complement formal surveys rather than replace them.

Table 1: The non-traditional data landscape at a glance

Data family	What it captures	Typical refresh	Spatial grain	Illustrative KPI uses
Satellite EO (optical, SAR, thermal, night lights)	Land use, buildings, vegetation, heat, floods, pollution columns, economic activity	Daily to weekly	0.3–20 m	Built-up expansion, green cover, heat island, flood extent
Drone / UAV imagery	Asset condition, vegetation health, 3D structure, compliance	On demand	Centimetres	Park health, façade condition, construction progress
Mobile network / location data	Dynamic population, flows, footfall, commuting	Hourly or finer	150–500 m grid	15-minute city access, event crowds, transit demand

Data family	What it captures	Typical refresh	Spatial grain	Illustrative KPI uses
IoT sensors	Air, noise, water, energy, waste, parking, structures	Seconds to minutes	Point / network	Hyperlocal Air Quality Index (AQI), leakage, bin fill, bridge health
Connected vehicles & transit data	Speeds, positions, taps, incidents	Real time	Road link / stop	Congestion index, bus reliability, road safety risk
Street-level imagery / video	Counts, defects, incidents, streetscape quality	Continuous / periodic	Street segment	Pedestrian volumes, pothole detection, walkability
Administrative digital exhaust	Service requests, permits, meters, transactions	Continuous	Address / account	Resolution times, digital adoption, energy per capita
Social & web data	Sentiment, complaints, prices, perception	Continuous	Variable	Citizen satisfaction proxy, emerging-issue detection

3 The AI Toolkit: From Raw Signals to Trusted Indicators

None of these data sources speaks the language of KPIs natively. A satellite image is millions of pixels; a month of mobility data is billions of events. Artificial intelligence is the translation layer.

In brief for city leaders. You do not need to master these techniques. You do need to know what each one is used for, and to insist that every AI-derived figure comes with a stated level of accuracy. Table 2 gives a one-page summary; the sub-sections that follow are written mainly for technical teams.

3.1 Computer vision and deep learning

Convolutional neural networks, and increasingly transformer-based and hybrid architectures, are the workhorses of imagery-based measurement. Semantic segmentation classifies every pixel of a satellite or aerial image into categories such as building, road, vegetation and water, which enables automated mapping of land use, green cover and impervious surfaces.

Object detection locates and counts discrete things such as vehicles, pedestrians, rooftop solar panels and damaged assets in video and imagery. Change detection compares images across time

to quantify construction, demolition, encroachment and disaster damage, so that a city can track built-up expansion across its metropolitan area in near-real time.

3.2 Supervised machine learning for estimation and inference

Gradient-boosting methods, random forests and neural networks translate behavioural traces into labels and estimates that no sensor directly measures. Examples include inferring transport mode (walking, cycling, bus, car or metro) from mobile location sequences; estimating fine-grained population from network events; predicting building energy intensity from characteristics and weather; and calibrating low-cost sensors against reference stations, where neural network calibration has materially improved the accuracy of city-wide networks. These models are also the standard tool for nowcasting official statistics: producing a timely AI estimate of an indicator whose official value arrives much later.

3.3 Time-series forecasting and predictive analytics

Forecasting models give KPIs a forward-looking dimension. The main techniques are recurrent neural networks (RNN), temporal convolutional models, gradient-boosted trees and, increasingly, graph neural networks (GNN) that respect the road-network topology. Pollution levels can be forecast hours ahead so that health advisories precede the spike rather than follow it. Congestion can be predicted before it forms. Water-main failure probabilities can be ranked so crews are dispatched before a pipe fails. This is the difference between reporting a problem and preventing one.

3.4 Natural language processing (NLP) and large language models (LLMs)

A large share of a city's self-knowledge is trapped in text: service requests, inspection notes, permit descriptions, complaints, social posts and news. NLP classifies and routes service requests automatically, measures sentiment as a continuous satisfaction proxy, clusters emerging issues before they trend, and extracts structured indicators from unstructured documents. Large language models add a conversational layer, letting a mayor ask a question of the KPI platform in plain language, and they speed up the analyst work of summarizing, comparing and explaining indicators.

3.5 Data fusion and multimodal learning

The most accurate urban indicators rarely come from a single source. Fusion techniques from classical Kalman filtering and spatial statistics to multimodal deep learning blend satellite, sensor, mobility and administrative data into one coherent estimate with quantified uncertainty.

3.6 Edge AI and real-time processing

Running models on or near the device that collects the data (the camera, the pole or the vehicle) rather than in a distant data centre cuts latency, bandwidth and privacy risk simultaneously. Edge AI enables intersection cameras that transmit only counts and incident flags, never faces or plates; acoustic sensors that report an event classification, not audio; and city-scale systems that keep functioning when connectivity fails. It is the architectural pattern that makes continuous, privacy-respecting measurement affordable at scale.

3.7 Digital twins and simulation

A digital twin binds all of the above into a live, navigable model of the city, continuously updated from IoT and imagery feeds. Twins do two things ordinary dashboards cannot. First, they contextualize KPIs on the map, showing not just that air quality declined but in exactly which streets, and why. Second, they make KPIs forward-looking: planners can test a proposed measure in the model and see its likely effect on the indicators before committing funds. The twin turns the KPI system from a scoreboard into a flight simulator.

3.8 Generative and agentic AI

Generative models and AI agents extend measurement into imagination and autonomy: generating realistic scenarios of urban growth from historical satellite series, synthesizing privacy-safe mobility datasets that preserve statistical properties without exposing individuals, drafting KPI narratives and briefings automatically, and orchestrating multistep analytical workflows. Cities are adopting these capabilities cautiously, keeping humans firmly in the decision loop, a stance codified in emerging municipal AI governance frameworks.

Table 2: The AI toolkit mapped to its role in KPI production

AI technique	What it does for KPIs	Representative urban application
Semantic segmentation & change detection	Pixel-level mapping of land, assets and change over time	Green-cover %, built-up expansion rate, disaster damage extent
Object detection & tracking	Counts and classifies things in imagery and video	Traffic volumes by mode, parking occupancy, defect detection
Gradient boosting / classic ML	Infers labels and quantities from behavioural traces	Transport-mode share from mobile data, sensor calibration
Forecasting models (RNN, GNN, boosted trees)	Predicts indicator values hours to weeks ahead	AQI forecasts, congestion prediction, asset-failure risk
NLP & large language models (LLMs)	Structures and interprets text at scale	311 auto-triage, citizen-sentiment index, plain-language KPI queries

AI technique	What it does for KPIs	Representative urban application
Multimodal data fusion	Combines sources into one calibrated estimate with uncertainty	Street-scale air quality (model + sensors), dynamic population
Edge AI	Real-time, privacy-preserving inference at the source	Camera counts without image retention, acoustic event detection
Digital twins & simulation	Live spatial context and what-if scenario testing	Planning-scenario evaluation, flood stress tests, fleet reliability
Generative & agentic AI	Scenario synthesis, privacy-safe data, automated analysis	Growth scenarios from satellite series, synthetic mobility data

4 From Data to Decision: The AI-Driven KPI Value Chain

An AI-driven KPI is not a model – it is a production line. Cities that succeed industrialize five stages and wrap them in governance.

In brief for city leaders. Producing a trusted AI-driven indicator takes five steps and three permanent safeguards. The hardest parts are usually obtaining, combining and checking the data, not building the AI models.

1. Collect the data. The first stage is acquisition and ingestion: establishing reliable pipelines from satellites, sensor networks, operators, platforms and internal systems, with clear legal bases, data-sharing agreements and quality checks at the point of entry.

2. Prepare and combine it. The second is preparation and integration: cleaning, de-duplication, anonymization and aggregation of personal data, geospatial alignment onto common grids and registers, and harmonization of legacy formats.

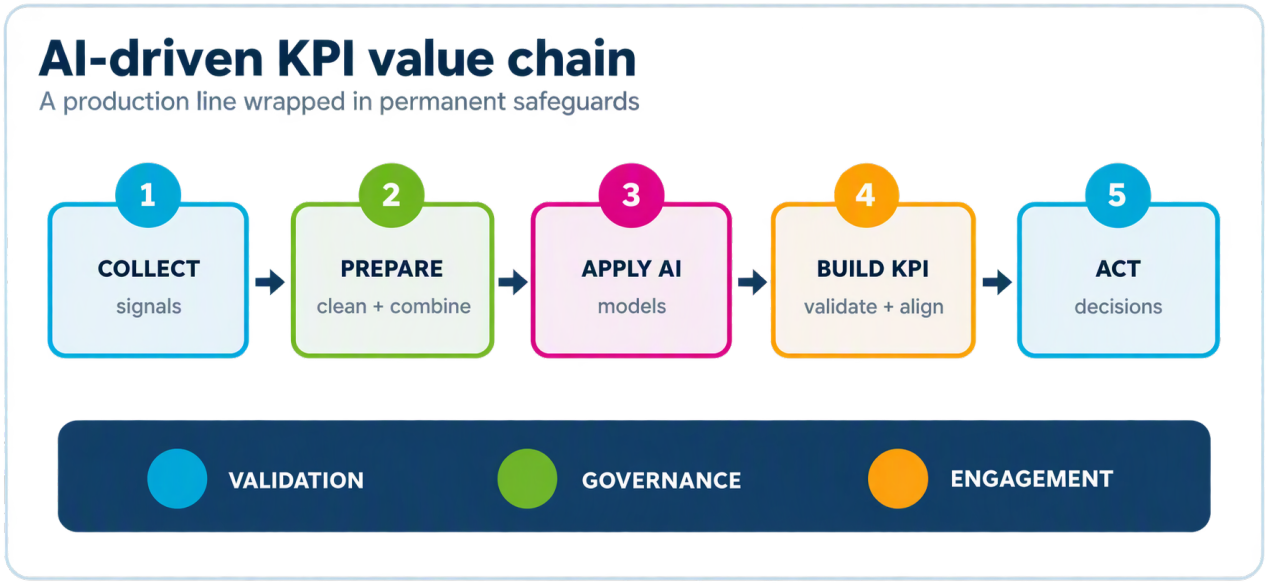
3. Apply the models. The third stage is intelligence: the machine-learning models of Section 3, trained, validated against ground truth, versioned and monitored for drift like any other critical infrastructure.

4. Build the indicators. The fourth is indicator assembly: converting model outputs into defined KPIs with documented methodologies, uncertainty ranges and alignment to frameworks such as the U4SSC KPIs and the UN-Habitat Global Urban Monitoring Framework, so that an AI-derived figure can sit credibly beside an official statistic.

5. Deliver and act. The fifth is delivery and action: dashboards, open-data portals, alerts, briefings and API feeds into operational systems, closing the loop so a KPI does not merely inform a report but triggers a signal-timing change, a maintenance order or a health advisory.

Around the pipeline sit three permanent functions:

Figure 4: The AI-driven KPI value chain and its three permanent safeguards



- a validation function that continuously compares AI estimates with reference measurements and publishes accuracy metrics;
- a governance function that manages privacy, ethics, procurement and model risk; and
- an engagement function that explains methods to the public and to auditors.

Table 3 summarizes what this production model changes relative to traditional measurement.

Table 3: What changes when KPIs move from periodic collection to continuous computation

Dimension	Traditional measurement	AI-driven measurement
Frequency	Annual, quarterly, or per survey cycle	Continuous – minutes to days
Spatial resolution	City or district averages	Street, block or 100–500 m grid
Latency	Months to years after the fact	Near-real time; often predictive
Marginal cost per data point	High – fieldwork and manual processing	Low once pipelines are built; scales inexpensively
Coverage	Sampled locations, extrapolated	Dense across the territory

Dimension	Traditional measurement	AI-driven measurement
Direction of insight	Descriptive (what happened)	Descriptive, diagnostic and predictive (what is about to happen)
Comparability	Method varies by survey and vintage	Identical algorithm across space and time; auditable versions
Key risks	Staleness, small samples, cost-driven gaps	Model error, bias, privacy, drift – requiring active governance

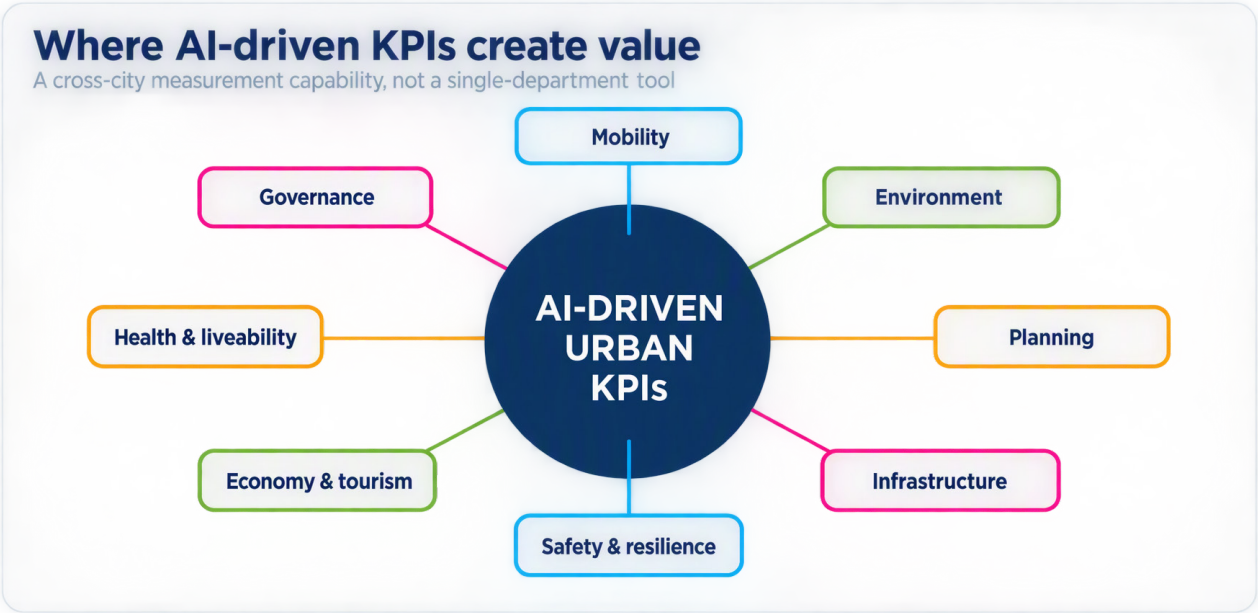
The two models are complements: traditional instruments become the calibration backbone of the AI system.

5 Where It Works: Application Domains Across the City

Various aspects of a city's scorecard can now be computed this way. What follows is a domain-by-domain tour, with illustrative KPIs and the data-plus-AI recipe behind each.

In brief for city leaders. Every major city department has indicators that can be produced this way. Mobility is the most mature domain. Table 4 lists sample indicators by domain, from which a city can choose its first two or three.

Figure 5: Application domains for AI-driven urban KPIs



5.1 Mobility and transportation

Mobility was the first domain to be transformed by non-traditional data, and it remains the most mature. Floating-car and connected-vehicle data produce continuous congestion indices and door-to-door travel-time KPIs for every road link; computer vision at intersections counts and classifies cars, buses, trucks, cyclists and pedestrians around the clock; machine learning infers the modal split from anonymized mobile traces, replacing travel surveys that were conducted once a decade; and transit telemetry plus smart-card data yields punctuality, crowding and accessibility indicators by line, stop and hour.

Illustrative KPIs include: average intersection delay, congestion index by corridor and hour, public-transport on-time rate, modal share, kilometres of road at level-of-service, road-safety risk score by segment, average detection-to-response time for incidents.

5.2 Environment, climate and energy

Dense low-cost sensor networks, calibrated and fused with satellite observations and dispersion models, give cities hyperlocal air-quality indicators, pollutant concentrations by neighbourhood and hour, population-weighted exposure, and forecasts that trigger advisories before conditions deteriorate. Thermal satellite imagery quantifies the urban heat island and identifies the streets where shade and greening will save lives; segmentation of optical imagery tracks green cover and canopy percentage continuously; night-time lights and smart-meter data feed energy-intensity and efficiency KPIs; and machine learning on consumption, weather and building data estimates greenhouse-gas emissions at district level between official inventories.

Illustrative KPIs include: hyperlocal Air Quality Index (AQI) and exceedance hours, population-weighted PM2.5 exposure, heat-island intensity by district, green cover percentage, estimated district CO_2 emissions.

5.3 Urban planning, land use and housing

Satellite change detection has made master-plan monitoring continuous: built-up expansion rate, densification, land consumed per new resident and encroachment on protected zones can be computed from open imagery every few weeks. Computer vision on very-high-resolution and drone imagery tracks construction progress against permits and detects unauthorized development and informal settlement growth early. Mobile-derived dynamic population combined with points-of-interest and network analysis powers accessibility KPIs, the share of residents within fifteen minutes of daily needs.

Illustrative KPIs include: annual built-up expansion versus plan, share of growth within transit catchments, 15-minute-city accessibility score, construction-permit compliance rate, vacancy and land-utilization indices.

5.4 Infrastructure and utilities

Vehicle-mounted cameras and AI turn every municipal fleet journey into a road-condition survey, producing pavement-quality indices and automatic pothole and signage-defect detection; acoustic and pressure sensors with anomaly-detection models locate water leaks and estimate non-revenue water continuously; smart meters feed consumption-efficiency KPIs; and structural-health sensors on bridges convert vibration signatures into condition scores. The frontier is predictive maintenance: ranking assets by failure probability so budgets flow where risk is highest.

Illustrative KPIs include: pavement condition index by segment, water losses as a share of supply, streetlight uptime, mean time from defect occurrence to detection, share of maintenance that is preventive rather than reactive.

5.5 Public safety, resilience and emergency management

In cities that deploy such systems, computer vision and acoustic sensing detect incidents such as collisions, fires, crowd surges and gunshots within seconds, which sharply shortens the detection component of response-time KPIs. Satellite radar and optical imagery quantify flood extent and post-disaster damage within hours, replacing weeks of ground assessment.

Illustrative KPIs include: median emergency response time, incident detection latency, share of population in modelled flood-risk zones, time to complete damage assessment after an event, crowd-density safety index at major venues.

5.6 Economy, tourism and urban vibrancy

Where business statistics arrive with long lags, non-traditional data offers a live economic pulse: anonymized footfall measures retail vitality street by street; aggregated payment data tracks local consumption; night-time light intensity serves as a validated proxy for economic activity; scraped listings and registries capture business openings, closures and rents; and mobile data quantifies tourism volumes, origins, stay duration and dispersal beyond the traditional hotspots.

Illustrative KPIs include: footfall index by high street, night-time economy activity index, business birth and survival rates, tourism-dispersal ratio, average visitor dwell time in cultural districts.

5.7 Health, wellbeing and liveability

Combining pollution surfaces with dynamic population produces exposure KPIs that reflect where people actually are, not just where they sleep; street-view imagery scored by computer vision yields systematic walkability, greenery and shade indices for every street; mobility traces measure active-travel minutes; and heat-vulnerability indices fuse thermal imagery with demographic and health data to target cooling interventions. These indicators connect the physical city to human outcomes and give health equity a measurable, mappable form.

Illustrative KPIs include: population-weighted pollution exposure, share of residents within a five-minute walk of quality green space, street-level walkability score, heat-vulnerability index, active-mobility minutes per capita.

5.8 Governance and service delivery

The city's own operational records measure its responsiveness. NLP auto-classifies and routes service requests and computes resolution-time KPIs by topic and neighbourhood; sentiment analysis of requests, reviews and social posts provides a continuous satisfaction proxy between formal surveys; telemetry from digital services yields adoption, completion and failure rates; and hotspot models predict where demand for services will surge.

Illustrative KPIs include: median service-request resolution time, first-contact resolution rate, digital-service completion rate, citizen-sentiment index, predicted versus actual demand accuracy.

Table 4: Consolidated sample KPI catalogue by domain – a menu from which aspiring cities can select their first lighthouse indicators

Domain	Signature AI-driven KPIs	Primary data	Core AI techniques
Mobility & transport	Congestion index, modal share, transit reliability, incident response time	Connected vehicles, video, mobile traces, smart cards	Computer vision, GNN forecasting, mode inference
Environment & energy	Hyperlocal AQI, heat-island intensity, green cover, district CO2 estimate	IoT sensors, satellites, smart meters	Sensor fusion, calibration ML, forecasting, segmentation
Planning & land use	Built-up expansion, 15-minute accessibility, permit compliance	Satellite & drone imagery, mobile data, Point of Interest (POI)	Change detection, segmentation, accessibility modelling
Infrastructure & utilities	Pavement index, water losses, streetlight uptime, preventive-maintenance share	Vehicle cameras, acoustic & pressure sensors, meters	Defect detection, anomaly detection, predictive maintenance
Safety & resilience	Response time, detection latency, flood exposure, damage-assessment time	Video, acoustic, SAR satellite, mobility data	Incident detection, flood modelling, rapid damage mapping
Economy & vibrancy	Footfall index, business dynamism, tourism dispersal, night-economy index	Mobile footfall, payments, night lights, web data	Nowcasting, NLP, econometric ML
Health & liveability	Exposure per capita, walkability score, heat vulnerability, green access	Street imagery, thermal satellite, mobility, sensors	CV streetscape audit, exposure fusion, risk indices
Governance & services	Resolution time, sentiment index, digital completion rate, demand forecast	311/CRM (Customer Relationship Management), service telemetry, social data	NLP classification, sentiment analysis, LLM assistance

6 From Concept to Practice: Case Studies

Three implementations show what AI-driven measurement delivers today, and where its limits still lie.

In brief for city leaders. Three working examples show measured gains in speed, accuracy or local detail. They also show that a legal mandate, clear ownership and published accuracy checks matter more than the choice of algorithm.

The three cases in this section are drawn from case studies prepared for the U4SSC Working Group on AI-Driven KPIs Measurement. Each has run on real data and reports measured results. Each is assessed against the three questions that guide the Working Group: does AI make the indicator faster to produce, more accurate, and more sensitive to differences between places and groups? Table 5 summarizes the cases. A box at the end of the section describes a complementary method for checking whether a city's data can support an indicator before measurement begins. Fuller accounts and source references are given in the individual U4SSC case studies.

Table 5: The three case studies at a glance

Case	Indicator	Data and AI	Reported result	Main limitation
Dubai Population Now (United Arab Emirates)	Resident population and daytime presence	Administrative registers from more than 12 entities; machine learning for data preparation, classification and forecasting	End-2024 population revised upward by about 10 per cent; continuously updated population clock; breakdown to more than 200 communities	Algorithms and detailed methodology not published; no independent evaluation yet
D-MOSS dengue forecasting (Viet Nam)	Forecast dengue incidence per 100 000 and probability of exceeding an outbreak threshold	Disease surveillance, climate, Earth observation, hydrological and population data; probabilistic ensemble forecasting	Forecast error 17.8 per cent lower than a seasonal baseline over 39 months; peak timing error of 2.5 months against more than four	Provincial rather than city scale; use in city decisions not publicly recorded
Project Sidewalk (35 cities, platform-reported)	Location and severity of pedestrian accessibility barriers	Street-level imagery and volunteer labels; computer vision and vision–language models	75–100 per cent agreement with government field audits on the presence of features	Agreement on severity as low as 20.8 per cent; not yet used in official reporting

6.1 Population: Dubai Population Now (United Arab Emirates)

Population is the denominator of most city KPIs, including the many U4SSC indicators expressed per inhabitant or per 100 000 inhabitants. Dubai conducted seven conventional censuses between 1975 and 2005 and has relied on estimates since. Dubai Population Now, led by the Dubai Data and Statistics Establishment within the Dubai Digital Authority, replaces periodic enumeration with a register-based approach. It links administrative data from more than 12 government and private entities, covering residency, transport, utilities, health, property, education and tourism. The Executive Council of Dubai approved the initiative on 1 July 2026, and the first results were released on 30 July 2026.

Machine learning supports the preparation of the linked records: data cleansing, deduplication, entity resolution and statistical classification. Predictive models are used to forecast demographic trends and to update a public population clock. The specific algorithms have not been published.

The measured effects fall along three dimensions. On accuracy, the register-based results led to a revision of the end-2024 population from 3 863 600 to 4 248 200, a difference of 384 600 or about 10 per cent, which changes every per-capita indicator computed from it. On speed, the reference figure is now a continuously updated clock, which read 4 741 335 on 30 July 2026 against 4 580 300 at the end of 2025; its monthly series showed a decline of about 61 000 in March 2026 followed by a recovery, a movement that an annual estimate would not reveal. On sensitivity, the release separates usual residents from daytime presence (estimated at 6 392 300) and breaks the population down by gender, five-year age group, nine planning sectors and more than 200 communities.

Two points qualify these results. The accuracy gain comes mainly from linking registers under a clear legal mandate, not from the algorithms, so investment in identifiers, register quality and data-sharing law comes before investment in models. In addition, no independent evaluation has yet been published, and the residency rules, the computation of daytime presence and the uncertainty of the estimates are not documented in detail. Personal data are protected by a statistical-confidentiality regime under which only aggregate statistics are published. The approach is transferable to cities that have a unique personal identifier, registers of adequate coverage, a legal mandate for the statistical authority and a secure data platform.

6.2 Health: dengue forecasting in Ho Chi Minh City and Ha Noi (Viet Nam)

The Dengue forecasting MOdel Satellite-based System (D-MOSS) is an early warning platform that turns a health indicator from a record of past cases into a forecast. It combines confirmed dengue cases with a seasonal climate ensemble, Earth observation data, hydrological model outputs and population denominators to produce probabilistic forecasts one to six months ahead, refitted every month. The system was funded through the UK Space Agency and developed by an international consortium with Viet Nam's National Institute of Hygiene and Epidemiology as the national lead. It has been included as a tool in national dengue control guidelines since November 2020. Its engine is an ensemble of probabilistic statistical models; it is not a deep learning system.

The case proposes three indicators: forecast dengue incidence per 100 000 population, the probability that incidence will exceed a locally chosen outbreak threshold, and the lead time at which the forecast is issued. Dengue incidence is not one of the standard U4SSC KPIs, so these are presented as a possible extension of the health indicators, not as adopted measures.

The evidence is a prospective evaluation of forecasts issued during live operation from July 2019 to September 2022, covering 692 617 reported cases. Forecast error was 37.2 per cent lower than a random comparator and 17.8 per cent lower than a seasonal average baseline. The error in predicting the timing of the epidemic peak was 2.5 months, against more than four months for both baselines. The ability to separate outbreak from non-outbreak months varied across four decision scenarios, with an area under the curve between 0.586 and 0.851 (on a scale where 0.5 is no better than chance and 1.0 is perfect). Performance was weaker in low-incidence northern areas, and the unusually early outbreak of June 2022 was not captured well at longer lead times.

The institutional picture is less complete than the scientific one. Forecasts are produced at provincial level, and the published results for Ho Chi Minh City refer to its boundaries before the administrative reform of July 2025, so thresholds and denominators need to be rebased for the enlarged city. Public sources do not record a case in which a forecast changed the timing or scale of a city's dengue response, nor do they identify a permanent service owner or operating budget. The lessons for other cities are to keep uncertainty inside the indicator, to set outbreak thresholds jointly with health authorities, and to record whether each forecast led to a decision.

6.3 Mobility and accessibility: Project Sidewalk

Project Sidewalk is an open-source platform developed by the Makeability Lab at the University of Washington. Volunteers move through street-level imagery and label accessibility features and barriers such as curb ramps, missing sidewalks, obstacles and surface problems, rating each on a five-point severity scale. AI supports the volunteers in three ways: computer-vision classifiers validate their labels and filter false positives, vision–language models generate guidance adapted to the type of street being audited, and clustering merges overlapping labels into single locations. The U4SSC pedestrian infrastructure indicator measures pedestrian zones as a share of city area, which does not show whether that area can be used by a person in a wheelchair. The platform changes the unit of measurement from area to individual barrier.

The platform lists 35 deployment cities. In its first deployment, in Washington, D.C., 797 online contributors placed 205 385 labels and audited the entire street network of about 1 730 km. In Zurich in 2024, a partnership between the University of Zurich and the city produced more than 9 000 data points for the city centre, which were published on the city's open-data portal. In Chandigarh in 2026, an audit of about 40 km of roads around 230 points of interest found that 1 644 of 2 913 assessed locations would benefit from accessibility improvements.

A 2025 peer-reviewed study compared platform data with government field audits in Seattle and in DuPage County, Illinois. Agreement on whether a feature exists was high, ranging from 75 per cent for pedestrian signals to between 98.7 and 100 per cent for missing sidewalks, and between 89.9 and 93.5 per cent for curb ramps. Agreement on the condition of curb ramps was much weaker: 63.8 per cent in Seattle and 20.8 per cent in DuPage County. The authors attribute the gap mainly to the time difference between the imagery and the field audits, and to the lack of standard criteria for rating severity.

Three limits follow. No government has yet been documented as using the data in statutory planning or official indicator reporting. Coverage depends on where volunteers are active and on the recency and extent of proprietary street imagery, so some neighbourhoods and streets remain unaudited. Condition ratings are not yet reliable enough to carry the same weight as presence data. Cities adopting the method should also expect to adapt the audit rules to local infrastructure norms and law, as the Chandigarh deployment did.

6.4 What the cases show

Read together, the cases support the argument of this document and add four practical qualifications.

- **Gains are real but uneven**. Each case improves at least one of speed, accuracy and sensitivity with measured evidence. None improves all three without qualification: Dubai's methodology is not yet public, dengue forecasts are weaker in low-incidence areas, and virtual audits are reliable for presence but not for condition.
- **Institutions matter more than algorithms**. Dubai's accuracy gain rests on a legal mandate to link registers. The dengue system's main open question is who owns and funds it. Project Sidewalk depends on local partners to recruit and sustain volunteers.
- **Published validation is what earns trust**. The dengue and pedestrian cases report where the method fails as well as where it works, which gives cities a factual basis for deciding how far to rely on the results. This is the ground-truth loop described in Section 8.3.
- **Official adoption lags technical capability**. Only the Dubai figures are official statistics. In the other two cases the indicator is technically available but is not yet part of statutory reporting or a recorded decision process.

Box 1: Before measuring: checking data readiness (Bologna, Italy)

A fourth contribution to the Working Group addresses an earlier step: establishing whether a city's existing data can support an indicator at all. It is a methodology proposal demonstrated on real data. It does not measure an indicator, and it has not been formally adopted by the Municipality of Bologna.

The method uses large language models to read the descriptions and metadata of the datasets on a city's open-data portal and to compare them with the official definition of an indicator. Applied to Bologna and to SDG indicator 11.2.1 on convenient access to public transport, it reduced 701 datasets to 270 distinct units and 49 thematic clusters, of which 10 clusters containing 91 datasets were relevant. Each judgement is recorded as a claim, the evidence for it and the reasoning, so that a reviewer can trace and contest it.

The assessment found that the core calculation of the indicator can be made with Bologna's open data, but that nine aspects are not covered. Four are missing entirely: disaggregation by persons with disabilities, the physical accessibility and comfort of stops, access to real-time service information, and survey data on travel behaviour. For each gap the method suggests the kind of data source that could close it, such as a household survey, a field audit or new sensing.

For cities, the value lies in the inventory step of the foundation phase described in Section 9.2: the output is a prioritized list of data gaps. The results have so far been checked by manual sampling only. Validation with a second model and replication on other cities and indicators are planned.

7 The Value Proposition: Why This Matters

In brief for city leaders. The case for investing rests on four gains: faster information, a fairer picture of every neighbourhood, a lower cost per measurement, and figures that are consistent and can be audited.

Faster information. Timeliness is the most visible gain. Indicators that once arrived annually now refresh hourly and increasingly arrive before the event: the pollution episode is forecast, the congestion is predicted, the asset failure is ranked by probability. Measurement stops being an audit of the past and becomes an instrument of anticipation. In operational terms, this is the difference between publishing a report and preventing a problem.

A fairer picture. Local detail (granularity) is the equity gain. City averages flatter the well-served and hide the neglected; block-level indicators expose exactly which neighbourhoods breathe worse air, wait longer for buses, walk hotter streets and receive slower services. Because sources such as satellites observe everywhere with identical methodology, AI-driven KPIs also reach informal settlements and peripheral districts that household surveys chronically under-sample, provided cities remain alert to biases in the data (see Section 8.2).

Lower cost per measurement. Economics is the scaling gain. Traditional collection has a linear cost curve: twice the measurement, roughly twice the cost. AI-driven measurement is front-loaded: the investment goes into pipelines, sensors and models, after which the cost of another indicator, another district or another refresh is small. The systems, data licenses and skilled staff still need sustained funding. Open satellite data, shared platforms and the falling cost of sensors keep lowering the entry price.

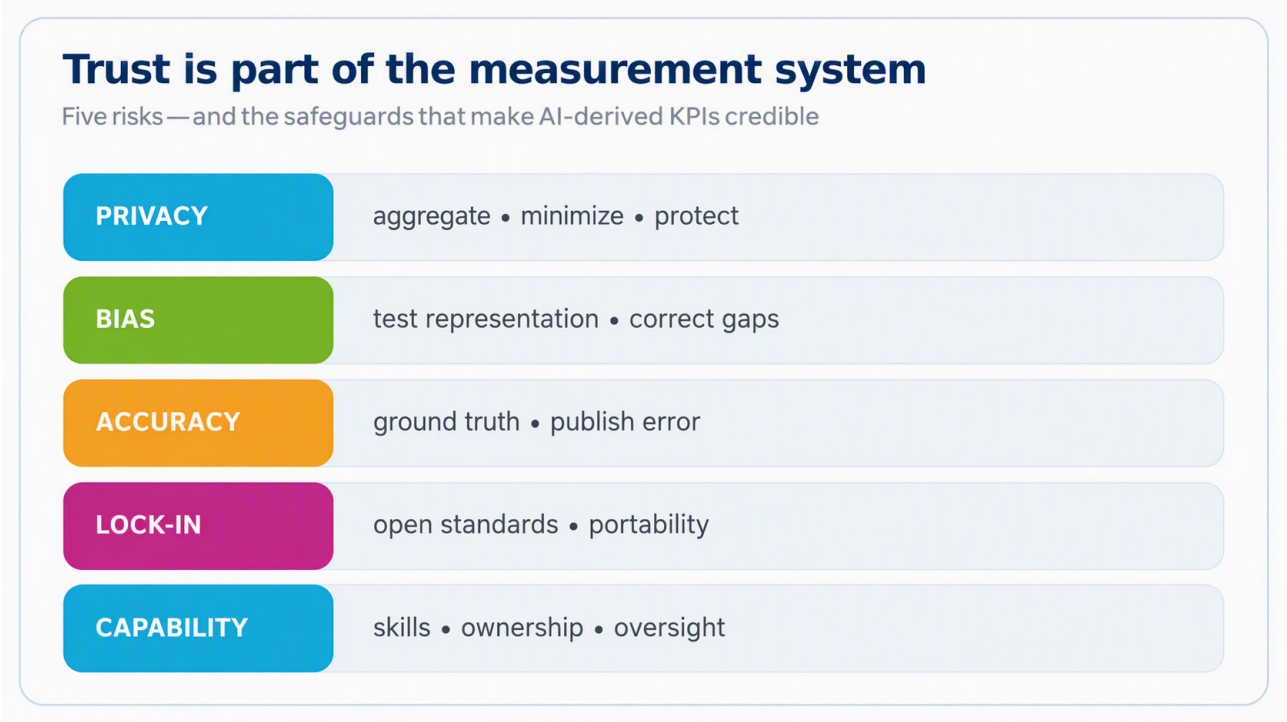
Consistency. Finally, there is a governance gain that is easy to underestimate: consistency. An algorithm applied identically across the whole territory, with versioned code and published accuracy, produces indicators that are comparable across space and time and auditable in a way that manually assembled statistics rarely are. Combined with alignment to international frameworks such as U4SSC KPIs, this positions AI-driven KPIs to feed official reporting, investor disclosure and inter-city benchmarking rather than existing as a parallel, unofficial universe.

8 Navigating the Challenges: Risks and Safeguards

Every capability in this document can be misused or mis-trusted. The cities that will lead this field are those that treat safeguards as part of the product, not an afterthought.

In brief for city leaders. Five risks need active management: privacy, bias, accuracy, vendor lock-in and skills. Each has known safeguards, set out below. A programme that loses public trust loses its data, so openness about methods and limits is essential.

Figure 6: Core risk areas and corresponding safeguards for trusted AI-driven KPIs



8.1 Privacy and social licence

Mobile traces, video and service records are personal at their origin even when indicators are aggregate at their destination. The mitigations are now well established: aggregate early and at the source, apply recognized anonymization techniques (such as k-anonymity thresholds and differential privacy) so that no individual can be singled out, prefer edge processing that transmits counts rather than images, use privacy-preserving synthetic datasets where individual-level analysis is needed, and subject every new data source to a documented privacy impact assessment. Just as important is transparency: publishing what is collected, why, and with what safeguards, because a KPI programme that loses public trust loses its data supply.

8.2 Bias and representativeness

Non-traditional data inherit the biases of the systems that generate them. Mobile data under-represent children, the elderly and those without smartphones; social media skew young and vocal; sensor networks cluster where installation was easy; and satellite classifiers trained on one urban form can misread another. The safeguard is statistical rigour: reweight against census benchmarks, validate across demographic segments, blend multiple sources so their skews offset, and publish known limitations alongside every indicator. Bias auditing should be scheduled and routine, not triggered by controversy.

8.3 Accuracy, validation and model drift

Low-cost sensors drift; models decay as the city changes; imagery misleads under clouds, shadow and seasonal variation. Credible programmes therefore maintain a permanent ground-truth loop, in which reference-grade stations, manual counts and survey waves are used to calibrate and score the AI estimates, and they publish accuracy metrics with the same discipline as the indicators themselves. Machine-learning calibration of low-cost networks and continuous benchmarking against official statistics are what elevate an interesting model into a publishable statistic.

8.4 Interoperability, legacy systems and vendor lock-in

The hardest engineering problems are rarely the models; they are the connections between systems. Aspiring cities should insist on open standards and APIs, data portability, documented schemas and exit clauses, and favour modular architectures over monolithic platforms, so that today's systems do not become tomorrow's strategic constraint.

8.5 Skills, institutions and regulation

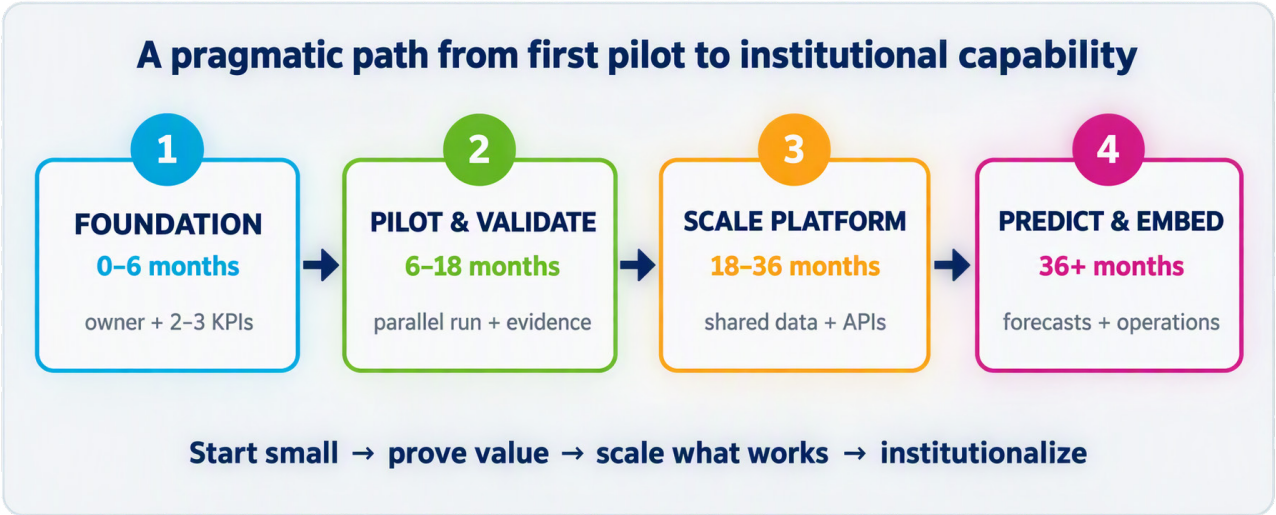
An AI-driven KPI programme is an institutional capability, not a one-off purchase. It needs data engineers, ML practitioners, geospatial analysts, privacy officers and, critically, domain experts who can challenge model outputs. Regulation is converging on risk-based oversight. Cities that build governance early will find compliance a formality; those that retrofit it will find it costly and difficult.

9 A Roadmap for Aspiring Cities

The destination is a continuously measured city. The journey is incremental, and the first steps are cheaper than most cities assume.

In brief for city leaders. Start small. Name an accountable lead, choose two or three indicators tied to real decisions, run the new method alongside the old one and publish the comparison. A committed city can reach level 3 of the maturity model within about two years.

Figure 7: Phased implementation path for aspiring cities



9.1 A maturity model for AI-driven measurement

Most cities can locate themselves on a five-level maturity curve. The purpose of the model is not to rank cities but to sequence investment: each level creates the data, skills and trust that the next one requires.

Table 6: Five levels of measurement maturity. Levels 1–3 are achievable by any committed city within two years; levels 4–5 describe the current global frontier.

Level	Name	Defining characteristics	Typical KPI cadence
1	Traditional	Surveys, manual counts, siloed spreadsheets; KPIs assembled by hand for annual reports	Annual or slower
2	Instrumented	First sensor pilots and dashboards; open-data portal; descriptive analytics on single sources	Monthly to quarterly
3	Integrated	Shared urban data platform; several AI-computed KPIs in production with documented validation; privacy governance in place	Daily to weekly
4	Predictive	Multisource fusion; forecasting embedded in operations; digital twin for scenario testing; published accuracy metrics	Real time + forecast
5	Adaptive	Closed-loop optimization with human oversight; KPIs trigger automated operational responses; continuous audit and public transparency	Continuous, self-correcting

9.2 A phased implementation path

Phase 1 – Foundation (months 0–6). Secure an executive mandate and name an accountable owner, typically a chief data or digital officer. Select two or three "lighthouse" KPIs (flagship indicators that demonstrate the approach) by starting from decisions, not data: choose indicators a named decision-maker will act on monthly. Inventory existing data assets and gaps, establish the privacy and ethics baseline, and deliver an early, visible result using free resources such as open satellite imagery.

Phase 2 – Pilot and validate (months 6–18). Build the first production pipelines for the lighthouse KPIs, deliberately spanning two different data families (for example, satellite plus IoT, or video plus mobility data) so that the city gains experience in combining sources. Run AI estimates in parallel with traditional measurement, publish the comparison honestly, and complete privacy impact assessments and a public methodology note. Partnerships accelerate this phase: telecom operators for aggregated mobility data, universities for model development, space agencies and international programmes for imagery-derived baselines.

Phase 3 – Scale the platform (months 18–36). Consolidate pilots onto a shared urban data platform with common geospatial reference layers, an API catalogue and sound model-management (MLOps) practices such as model versioning, drift monitoring and scheduled revalidation. Expand the KPI portfolio across at least four domains from Section 5, align definitions with U4SSC KPIs, the UN-Habitat Global Urban Monitoring Framework and national frameworks, and launch a public KPI portal that shows uncertainty ranges, not just point values. Institutionalize the operating model: a small central data team, domain product owners and a standing ethics review.

Phase 4 – Predict and institutionalize (months 36 and beyond). Add the predictive layer (forecasts, early warnings and risk rankings) and connect KPIs to operations so that indicators trigger action within defined human-oversight rules. Stand up or extend a digital twin for scenario testing, contribute data and methods back to open programmes, and benchmark internationally. By this stage, the annual report is no longer the product of the measurement system; it is a by-product of it.

10 Conclusion: From Periodic Reports to a Self-Aware City

Cities do not suffer from a lack of data; they suffer from a lack of timely, granular and decision-ready insight. For decades, urban performance management has relied on indicators that arrive too late, at too coarse a scale and at too high a cost to guide day-to-day governance. In an era defined by rapid urban change, climate risk and heightened public expectations, this model is no longer sufficient. Governing with last year's averages is increasingly incompatible with governing complex, living systems.

AI-Driven Urban KPIs offer a credible and already emerging alternative. By applying artificial intelligence to non-traditional data from satellites, sensors, mobility systems, digital services and

the city's own operational exhaust, cities can measure themselves as they are, street by street and hour by hour, and increasingly as they are about to be. Indicators shift from static scorecards to living signals. Measurement moves from retrospective reporting to anticipatory capability. Performance management becomes not just descriptive, but predictive and preventive.

The impact of this shift is not merely technical. It is strategic. Near-real-time, neighbourhood-scale indicators make disparities visible where they actually occur and therefore governable. Predictive KPIs allow leaders to act upstream, reducing harm rather than documenting it. Once pipelines are in place, the economics of measurement change fundamentally: the additional cost of better insight falls sharply, allowing cities to expand what they measure without proportionally expanding budgets. Just as importantly, algorithmic consistency across space and time improves comparability, auditability and alignment with international frameworks, positioning AI-driven KPIs to complement and strengthen official statistics rather than compete with them.

Yet this document has argued throughout that capability without governance is liability. AI-driven KPIs are not shortcuts around statistical discipline, privacy or accountability. They are a new production method for public metrics, and they demand safeguards commensurate with their power. Privacy-by-design architectures, continuous validation against ground truth, bias auditing, transparent methodologies and sustained human oversight are not optional features; they are prerequisites for legitimacy. The cities that will lead are not those that deploy the most sensors or the most models, but those that industrialize trust alongside insight.

Crucially, the journey does not begin at the frontier. Most cities can take practical first steps with assets they already possess: open satellite imagery, existing sensor pilots, administrative data and a small number of decision-critical KPIs owned by named leaders. Progressively integrating sources, publishing validation results and institutionalizing governance builds the technical, organizational and social foundations required for more advanced capabilities. As the maturity model illustrates, continuous and predictive measurement is achievable incrementally, not only by global megacities but by municipalities of varied size and capacity.

The long-term implication is profound. A city with AI-driven KPIs woven into operations becomes, in effect, self-aware. It can sense stress before failure, inequality before unrest and opportunity before it passes. Annual reports do not disappear, but they become summaries of a system that already knows itself in far greater detail. The role of leadership shifts accordingly: from commissioning measurements to acting on signals, from debating whose numbers are correct to deciding what to do next.

The choice facing city leaders is therefore not whether AI will shape how cities are measured; that transition is already underway. The choice is whether it will happen deliberately, transparently and in service of public value, or incrementally and opaquely through fragmented pilots and vendor solutions. Cities that act now, with ambition tempered by governance, can turn measurement into one of their most powerful levers for resilience, equity and performance in the decade ahead.


NEON CIT
START



For more information,
please contact: u4ssc@itu.int
Website: https://u4ssc.itu.int/

ISBN 978-92-61-43741-1

Published in Switzerland
Geneva, 2026

Photo credits: @AdobeStock