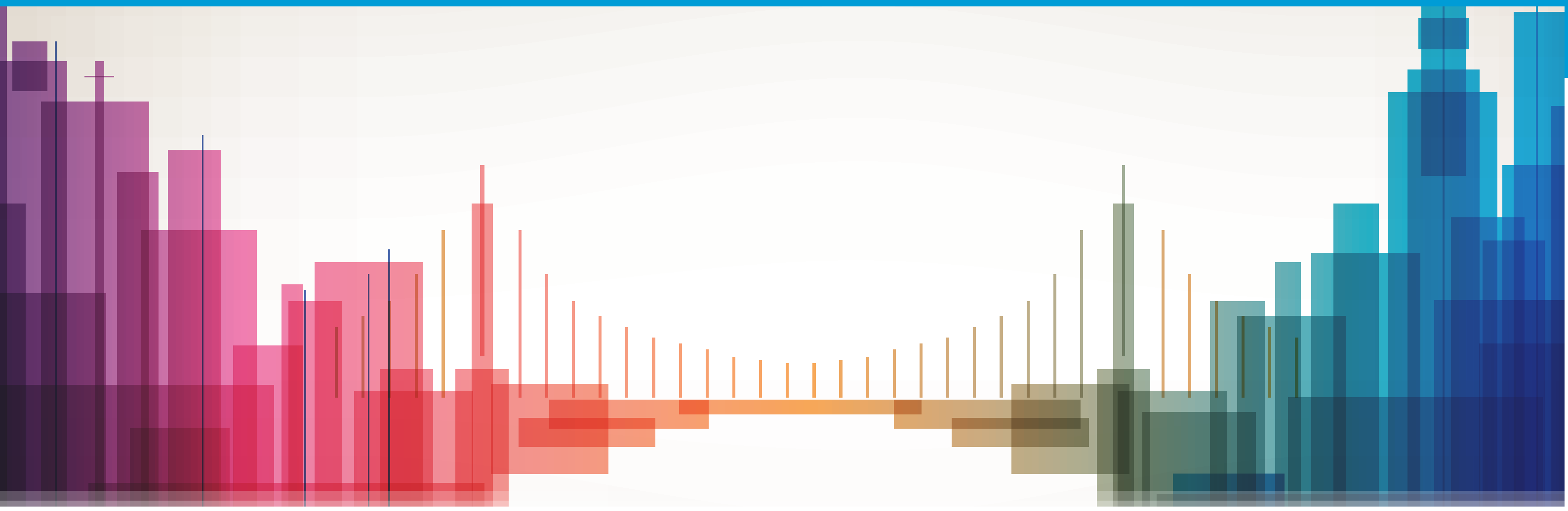
U4SSC

Autonomous cities and AI: The next frontier of urban transformation



Convention on Biological Diversity
United Nations ECLAC
Food and Agriculture Organization of the United Nations
UNIDO
ITU
UNDP
United Nations Economic Commission for Africa
UNECE
UNESCO – United Nations Educational, Scientific and Cultural Organization
UN environment programme
UNEP FINANCE INITIATIVE
United Nations Framework Convention on Climate Change
unitar – United Nations Institute for Training and Research
UN-HABITAT – FOR A BETTER URBAN FUTURE
United Nations – Department of Economic and Social Affairs
UN OFFICE 4 PARTNERSHIPS
UNU EGOV
UN Tourism
UN WOMEN
WORLD METEOROLOGICAL ORGANIZATION

Autonomous cities and AI: The next frontier of urban transformation

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 (Digital Dubai, United Arab Emirates).

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

The author would like to thank the following contributing individuals: Houssam Al Masri (Khatib & Alami, Saudi Arabia), Xiaomi An (Renmin University of China), Jose Esteban Gabarda (OSICO Platform), Guihong Chen (Beijing Big Data Centre), Zhangwen Lian (Shenzhen National Engineering Laboratory for E-Government Cloud Computing Application Technology Co., Ltd.), Zheng Qingguo (China Mobile Xiongan Information Communication Technologies Ltd., China), Gang Song (Command Centre of Beijing Municipal Bureau of Coordinated Administrative Law Enforcement for Urban Management), Val Wise (Independent Researcher), and KeWei Zhang (Kookmin University),

The author wishes to thank the U4SSC management team: Okan Geray (U4SSC Chair), 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 his 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), Simone Borelli 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), 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 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), Aline Matta, Edlam Abera Yemeru and Roberta Maio from the United Nations Human Settlements Programme (UN-Habitat), Dario Liguti, Tea Aulavuo from the United Nations Economic Commission for Europe (UNECE), Katarina Barunica Spoljaric and Nicholas Dehod from the United Nations Industrial Development Organization (UNIDO), 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), Sylvia Hordosch from the United Nations Entity for Gender Equality and the Empowerment of Women (UN-Women), 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.

ISBN

978-92-61-43001-6 (electronic version)
978-92-61-43011-5 (EPUB version)

Contents

List of figures and tables

Figures

Tables

Abbreviations and acronyms

Abbreviation	Full form
AI	Artificial intelligence
API	Application programming interface
HIC	Human-in-command
HITL	Human-in-the-loop
HOOTL	Human-out-of-the-loop
HOTL	Human-on-the-loop
ICT	Information and communications technology
IoT	Internet of Things
LiDAR	Light detection and ranging
LLM	Large language model
NLP	Natural language processing
SDG	Sustainable Development Goal
SDGs	Sustainable Development Goals
U4SSC	United for Smart Sustainable Cities

Executive summary

This guide presents a structured and comprehensive framework for understanding, assessing and progressively advancing autonomous cities. It defines autonomous cities as urban environments in which constituent systems such as mobility, energy, water, safety, health, public services and digital infrastructure operate with varying levels of sensing, decision-making, execution and learning capability. The concept is not intended to imply a city without human governance. Rather, it describes a continuum in which human institutions, public objectives and autonomous technologies interact to improve urban performance and resilience.

The report begins by distinguishing autonomous systems from automated systems. Automated systems execute pre-defined tasks with limited human intervention, whereas autonomous systems can perceive their environment, process information, make decisions and adapt behaviour within defined operational boundaries. Artificial intelligence, including machine learning, computer vision, natural language processing, large language models and agentic AI, is an important enabler of autonomy. However, autonomy also depends on sensors, actuators, control systems, data governance, operational procedures and institutional safeguards.

A central proposition of the guide is that cities are systems of systems. Urban life depends on the interaction of multiple infrastructures, services, institutions and communities. Changes in one system can produce ripple effects across others. For example, a disruption in the energy grid may affect transport, health services, digital connectivity and public safety. The autonomous city should, therefore, be understood as a layered and interdependent operating environment, not as a collection of isolated technology deployments.

The guide introduces a high-level model of autonomous systems built around five core components: sensors; sensing, perception and environment modelling; decision-making and action planning; actuators; and a learning module. This model provides a common language for assessing autonomy across diverse urban domains. It also supports the mapping of city systems into comparable capability dimensions, enabling city administrators to identify where autonomy is appropriate, where human oversight is essential, and where risk controls must be strengthened.

The assessment framework proposed in this guide contains two categories. First, component-specific assessment evaluates the extent of autonomy across perception, decision-making and learning capabilities. Second, general assessment examines cross-cutting requirements, including lawfulness, privacy, fairness, transparency, accountability, safety, security, robustness, sustainability, interoperability, accessibility, reliability, maintainability and human intervention. Together, these dimensions allow cities to assess not only whether a system is technologically autonomous, but also whether it is governable, trustworthy and inclusive, and fit for public use.

The guide also provides a practical four-step methodology: baseline the current extent of autonomy; determine the target autonomous state; implement actions to close identified gaps; and evaluate outcomes. This methodology is intentionally iterative. Cities may begin with a narrow set of systems, learn from implementation, and gradually expand the scope as institutional capacity, data infrastructure and public confidence mature.

The key conclusion is that autonomous cities should not be pursued as a single end-state or as a purely technical ambition. They should be treated as a governed transition towards more adaptive, responsive and sustainable urban systems. The responsible deployment of autonomous systems requires clear objectives, proportionate risk management, staged implementation, human-centred design, rigorous testing and continuous oversight. When applied carefully, autonomous systems can support better resource allocation, faster response to disruptions, improved service quality and more evidence-based urban governance.

Key Message	Policy Implication
Autonomy is a continuum, not a binary state.	Cities should define acceptable levels of autonomy by system, risk category and public value objective.
Cities are systems of systems.	Assessment must consider interdependencies and ripple effects across mobility, energy, water, digital infrastructure, health and services, etc.
AI enables autonomy but does not replace governance.	Human oversight, accountability, auditability and legal compliance must remain embedded in the operating model.
Autonomous city maturity varies across domains.	A city-level autonomy score should aggregate system-level scores using transparent and risk-sensitive weighting.
Responsible adoption requires iteration.	Cities should baseline, target, implement and evaluate in cycles, expanding scope as capability and trust mature.

1 Introduction to autonomous systems

Autonomous systems are self-governing entities capable of operating within a specified environment with a defined degree of independence. They sense conditions, process information, make decisions and act on those decisions in pursuit of stated objectives. Such systems may operate in physical environments, digital environments or hybrid cyber-physical settings. In urban contexts, they may support functions ranging from traffic optimization and predictive infrastructure maintenance to public safety monitoring, building management, robotic services and adaptive public administration.

For policy purposes, the term "autonomous system" should be interpreted carefully. Autonomy does not mean unlimited discretion, absence of regulation or removal of public accountability. It refers to a system's capacity to perform selected functions without direct human intervention, within boundaries established by law, policy, technical design and institutional governance. The degree of autonomy can, therefore, vary widely across systems and use cases.

The relevance of autonomous systems to city administrations is increasing as urban systems become more data-rich, digitally connected and operationally complex. However, their deployment raises important questions: which functions should be automated or autonomous; what level of human oversight is required; how accountability is assigned; how risk is assessed; and how public trust, safety, privacy, accessibility and inclusion are preserved.

1.1 Relationship to artificial intelligence (AI)

Artificial intelligence is a major enabling capability for autonomous systems. AI techniques allow systems to interpret data, identify patterns, forecast outcomes, support decision-making and adapt to changing conditions. Machine learning can improve system performance over time; computer vision can interpret visual environments; natural language processing can enable human-system interaction; and agentic AI can support goal-oriented planning and execution.

Nevertheless, autonomous systems should not be equated with AI alone. A system may include AI models but remain non-autonomous if decisions are made entirely by humans. Conversely, some systems may demonstrate limited autonomy through rules-based control, even without advanced AI. Autonomy emerges from the interaction of AI, data, sensors, actuators, control logic, operational rules, human oversight and institutional authority.

1.2 Relationship to automated systems

Automated and autonomous systems both reduce the need for manual execution, but they differ in the degree of independence and adaptability. Automated systems typically execute pre-defined tasks under predictable conditions, often following fixed rules. Examples include a washing machine cycle, an assembly-line robot or a conventional traffic signal programmed on a fixed schedule.

Autonomous systems, by contrast, are designed to operate in more dynamic environments. They may perceive real-time conditions, evaluate alternative actions, adjust behaviour and learn from experience. A self-driving vehicle, for instance, does not merely follow a fixed sequence; it interprets roads, vehicles, pedestrians, traffic signals and hazards, and adjusts its actions accordingly.

It is, therefore, useful to view automation and autonomy as part of a continuum of self-governance. At one end are human-operated systems; in the middle are systems that augment human decision-making; and at the higher end are systems capable of independent operation within defined constraints. This continuum is especially important for cities, where risk, public value and human oversight requirements differ across domains.

Dimension	Automated system	Autonomous system
Operating logic	Executes pre-defined rules or workflows.	Interprets conditions and may select actions dynamically.
Environment	Best suited to stable, predictable settings.	Can operate in changing and uncertain environments within defined boundaries.
Human role	Humans configure, supervise and intervene when needed.	Humans establish objectives, constraints, oversight and escalation mechanisms.
Learning capability	Often limited or absent.	May improve performance through data, feedback and model updates.
Urban example	Fixed-time traffic signal.	Adaptive traffic management system using real-time congestion and safety data.

1.3 Key characteristics of autonomous systems

Autonomous systems represent a paradigm shift in technology, pushing the boundaries of automation and intelligence. Some of the key characteristics of autonomous systems are listed below.

Characteristic	Policy-relevant description
Self-sensing	Ability to perceive the environment through physical or digital sensors, including cameras, LiDAR, radar, IoT devices, system logs and external data feeds.
Perception and interpretation	Ability to convert raw data into meaningful situational understanding, including object detection, event recognition, anomaly identification and environmental modelling.
Decision-making	Ability to determine a course of action using rules, optimization models, AI models, risk thresholds or learned strategies.
Action execution	Ability to execute decisions through physical actuators, digital commands, operational systems or service workflows.

Characteristic	Policy-relevant description
Adaptability	Ability to adjust behaviour in response to changing conditions, new data or feedback from past performance.
Human oversight	Ability to support, escalate to or be constrained by human decision-makers where required by law, risk or public interest.
Accountability and auditability	Ability to record actions, explain decisions and support review, correction and accountability.

Autonomous systems promise a confluence of cutting-edge technologies creating self-governing entities, where man-made machines handle complex tasks, all without constant human intervention.

1.4 AI as the engine that powers autonomy

AI can be understood as a key engine of autonomy because it enables systems to move from simple rule execution towards interpretation, prediction, optimization and adaptive decision-making. However, AI should be governed as part of a broader socio-technical system. The performance and legitimacy of autonomous systems depend not only on algorithms, but also on data quality, system integration, institutional readiness and the safeguards surrounding deployment.

Machine learning enables systems to learn from operational data and improve prediction, classification or optimization over time.

- Computer vision enables autonomous systems to interpret visual environments and identify objects, hazards, patterns and infrastructure conditions.
- Natural language processing enables interaction between humans and autonomous systems, including voice commands, service requests, explanations and reporting.
- Large language models can support complex interpretation and communication tasks, but should be governed carefully when used in public decision contexts.
- Agentic AI can support planning and task execution towards defined goals, but requires constraints, monitoring, accountability, and escalation pathways.

AI is, therefore, an enabling layer, not a substitute for governance. In city environments, AI-enabled autonomy must be embedded within lawful mandates, operational controls, cybersecurity standards, data governance arrangements and transparent accountability structures.

However, AI's role is not absolute. Autonomous systems additionally rely on:

- **Sensors:** These components function as the sensory mechanisms of the system, capturing and relaying critical environmental data such as temperature, pressure, air quality, and distance.
- **Actuators:** These elements convert computational decisions into physical actions, such as motor movements in robots, control surface adjustments in drones, and other urban devices and equipment that activate a certain functionality.

- **Control Systems:** These act as the main command centre, processing sensor data, executing AI algorithms, and issuing commands to actuators.

1.5 Challenges of autonomous systems

Autonomous systems introduce significant opportunities, but also new risks. These risks are amplified in urban settings because city systems are interconnected, affect large populations and often involve essential services. A failure in an autonomous system may have cascading consequences across infrastructure, safety, service delivery and public trust.

Challenge	Implications for city administrations
Safety	Systems must be tested against expected and edge-case scenarios, with fail-safe mechanisms and clear human override provisions.
Security	Autonomous systems may be vulnerable to cyberattacks, data manipulation, spoofing or unauthorized control. Security must be embedded across hardware, software, communications and operating procedures.
Ethics and public trust	Automated decisions may affect residents unequally. Cities must address fairness, inclusion, explainability and the right to contest or appeal where decisions affect people.
Legal accountability	Liability and responsibility must be clearly assigned across public authorities, vendors, operators, system designers and oversight bodies.
Sustainability	Autonomous systems may increase energy consumption, hardware dependency and e-waste unless lifecycle impacts are considered.
Readiness	Deployment requires data infrastructure, skills, procurement capability, governance frameworks, public communication and operational maturity.
Interoperability	Urban autonomy depends on systems working together. Common standards, APIs, data models and governance protocols are essential.
Human autonomy	Residents and public officials should retain meaningful agency, particularly in high-impact contexts involving rights, access to services or safety.

2 Autonomous cities

2.1 Cities as complex systems and systems of systems

Cities should be understood not merely as physical or built environments, but as complex, adaptive systems composed of multiple interconnected and interdependent components that collectively sustain and enable urban life. These systems are dynamic in nature, continuously evolving in response to demographic shifts, technological advancements, environmental pressures, and policy interventions.

A city, when conceptualized as a system, comprises several critical subsystems, including transportation networks, energy infrastructure, water and sanitation systems, communication frameworks, public services, and economic systems. Each subsystem performs distinct yet complementary functions, while remaining intrinsically linked to others through flows of resources, information and human activity. The functionality and stability of the city as a whole are contingent upon the reliability, coordination, and synchronization of these subsystems. For example, the effective operation of transportation signalling systems depends on an uninterrupted energy supply, while digital communication systems underpin the coordination of emergency services and governance functions.

Cities also operate through continuous flows of inputs and outputs. They draw essential resources such as food, water, energy, labour, and raw materials from surrounding regions and global supply chains. These inputs are transformed within the urban system into goods, services, economic outputs, and, inevitably, waste. The efficiency, sustainability, and circularity of these processes are critical determinants of urban resilience and long-term viability. Effective waste management, resource optimization, and sustainable consumption patterns are, therefore, central to maintaining the functional integrity of cities.

A defining characteristic of cities as systems is the presence of "emergent properties," whereby the interactions among subsystems generate outcomes that exceed the capabilities of individual components. Economic productivity, innovation ecosystems, social cohesion, and cultural identity are examples of such emergent phenomena. These properties arise not from isolated subsystems, but from the complex interplay between infrastructure, institutions, markets, and human behaviour. As such, they are inherently difficult to predict and require integrated, system-wide approaches to governance and planning.

Given these interdependencies, disruptions or inefficiencies in one subsystem can propagate across others, producing cascading or "ripple" effects throughout the urban system. For instance, disruptions in transportation infrastructure can affect workforce mobility, supply chains and economic productivity. Similarly, failures in water or energy systems can have immediate and far-reaching implications for public health, safety and overall quality of life. These cascading dynamics underscore the importance of adopting a holistic and anticipatory approach to urban management.

Importantly, cities are not singular systems but hierarchical "systems of systems." Each subsystem, such as transportation or energy, is itself composed of multiple interconnected subsystems and components. These nested relationships create layers of complexity, where interactions occur within and across different levels of the system. This hierarchical structure necessitates governance models that can operate across multiple scales, integrating local, city-wide, and regional considerations.

Adopting a systems perspective provides significant benefits for policymakers and city administrators. It enables more informed and data-driven planning, supports efficient resource allocation, and enhances the ability to anticipate and manage risks. Furthermore, it facilitates the identification of cross-sectoral synergies, enabling innovative solutions to complex urban challenges such as congestion, energy consumption, and environmental sustainability. Ultimately, recognizing cities as systems of systems is essential for advancing resilient, inclusive, and sustainable urban development.

2.2 Autonomous cities as a special case of autonomous systems

Building upon the conceptualization of cities as systems of systems, the notion of an "autonomous city" represents an advanced and evolving paradigm in urban governance and operations. In this context, cities integrate principles of autonomous systems, leveraging artificial intelligence (AI), advanced analytics, sensor networks, and digital infrastructure, to enable varying degrees of self-regulation, adaptive decision-making, and automated execution of urban functions.

An autonomous city may be defined as a self-governing urban system in which key infrastructures such as energy grids, water systems, transportation networks, and public services are increasingly optimized through real-time data collection, continuous monitoring, and algorithmic control mechanisms. These capabilities support enhanced efficiency, improved service delivery, and greater sustainability, while also enabling cities to respond dynamically to changing conditions and emerging challenges.

Core characteristics of autonomous cities include predictive maintenance, whereby embedded sensors and monitoring systems detect anomalies or potential failures in infrastructure and enable pre-emptive intervention before disruptions occur. Additionally, participatory AI mechanisms allow for the integration of citizen inputs, preferences, and behavioural data into planning and decision-making processes, thereby enhancing inclusivity and responsiveness. Adaptive regulatory frameworks further enable cities to dynamically adjust policies and operational rules in real time, for example by modifying traffic flows during peak periods or implementing environmental controls in response to air quality data.

More broadly, autonomous systems within cities are characterized by their capacity to sense their environment, process large volumes of data, generate insights, make decisions, and execute actions with limited direct human intervention. These capabilities are enabled by the convergence of technologies such as the Internet of Things (IoT), edge computing, cloud infrastructure, and advanced AI models. However, unlike narrowly defined autonomous systems such as self-driving vehicles with clearly bounded objectives, cities encompass complex socio-economic ecosystems with diverse stakeholders, competing priorities, and ethical considerations.

As such, human oversight, governance frameworks, and institutional accountability remain indispensable components of autonomous city models. Decision-making authority cannot be fully delegated to automated systems; rather, a hybrid approach is required, where human judgment and machine intelligence complement one another. This includes establishing clear governance mechanisms, ethical guidelines, and regulatory safeguards to ensure that autonomous systems operate in alignment with societal values, legal frameworks, and public interests.

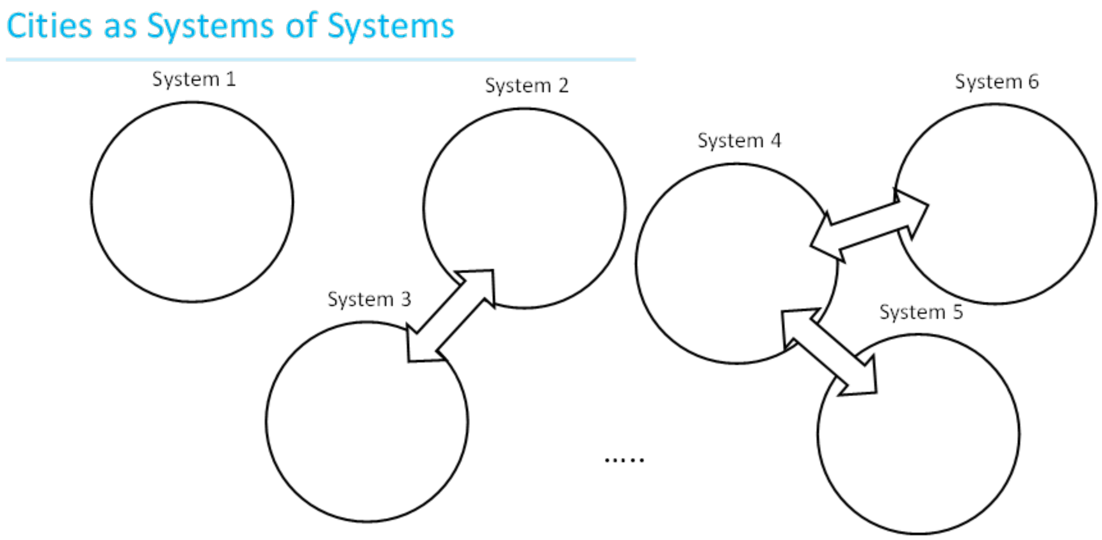
It is important to note that the concept of autonomous cities remains largely exploratory and is subject to significant technological, regulatory and societal challenges. These include issues related to data governance, cybersecurity, interoperability, public trust, and the equitable distribution of benefits. Nonetheless, the concept provides a forward-looking framework for understanding how cities may evolve towards greater autonomy, resilience, and sustainability through the integration of advanced digital and AI-driven systems.

In this regard, autonomous cities should not be viewed as an end state, but rather as a continuum of capabilities that can be progressively developed and implemented over time. This evolutionary perspective allows city administrations to adopt a phased and strategic approach, prioritizing high-impact use cases while ensuring that governance, infrastructure, and institutional readiness evolve in parallel.

3 Autonomous cities framework

The previous section has modelled cities as systems of systems. The figure below illustrates this concept.

Figure 1 – Cities composed of systems (Source: Author)



Each system represented in Figure 1 may comprise multiple subsystems, thereby forming hierarchical structures, as outlined in Section 2. This hierarchical composition reinforces the characterization of cities as systems of systems, wherein each layer consists of interdependent components operating at varying levels of granularity.

The directional linkages illustrated by the arrows in Figure 1 signify the functional and operational interdependencies between different urban systems. These linkages represent the transmission of effects, often referred to as "ripple effects", where changes or disruptions in one system propagate across others. Such interdependencies underscore the integrated nature of urban environments and highlight the necessity for coordinated planning and governance approaches.

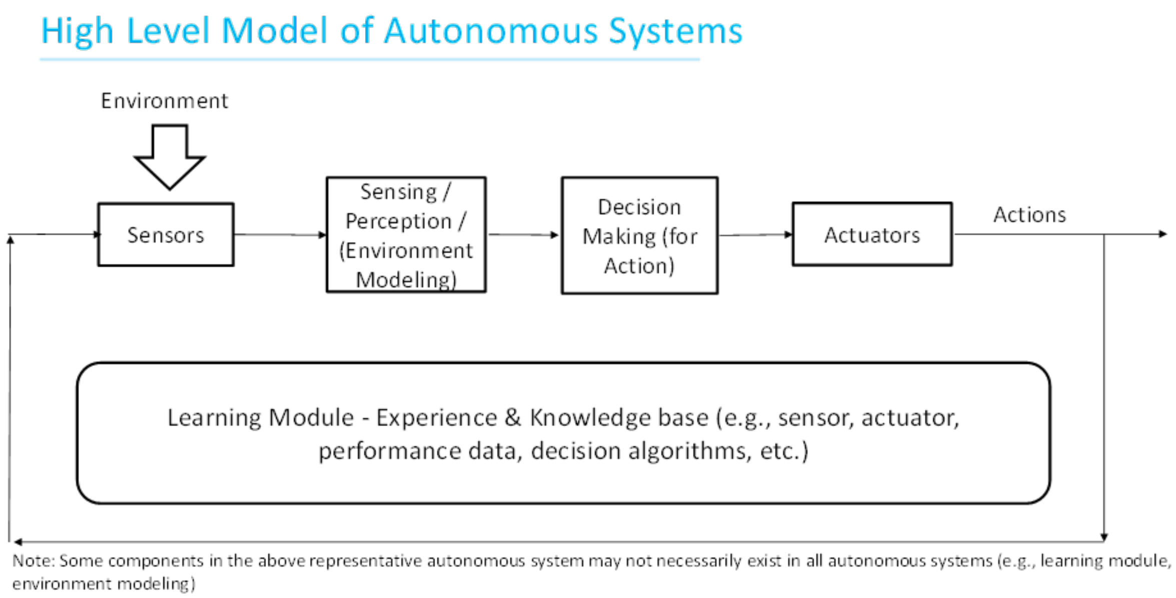
Accordingly, a city may be formally understood as a structured aggregation of interconnected and interdependent systems, each contributing to the overall functioning, resilience and sustainability of the urban environment.

3.1 High-level modelling of autonomous systems

As established in preceding sections, autonomous systems are characterized by their ability to perceive their environment, process and analyse data, make decisions, and execute actions with varying degrees of independence from human intervention. These systems are also capable of adapting to changing conditions and improving their performance over time through learning mechanisms.

Figure 2 presents a high-level conceptual model of an autonomous system, identifying its core functional components and their interactions. This model serves as a foundational framework for understanding how autonomy is operationalized within individual systems and broader urban contexts.

Figure 2 – Representative autonomous system (Source: Author)



Core components of autonomous systems

The principal components of the high-level autonomous system model are described below.

Sensors

Sensors constitute the primary data acquisition layer of the autonomous system, functioning as its "sensory interface" with the external environment. These components collect real-time data across multiple modalities, enabling situational awareness. Common examples include cameras for visual input, LiDAR systems for three-dimensional spatial mapping, radar for distance and velocity measurement, ultrasonic sensors for short-range detection, and environmental sensors for parameters such as pressure, temperature and humidity.

Sensing, perception and environment modelling

This component processes and interprets raw data obtained from sensors to generate a coherent and structured representation of the environment. It enables the identification, classification and tracking of objects and entities such as pedestrians, vehicles, infrastructure elements, and environmental conditions. Through advanced perception algorithms, the system develops a real-time situational model that informs subsequent decision-making processes.

Decision-making and action planning

The decision-making module functions as the central processing unit or "cognitive core" of the autonomous system. It analyses the interpreted environmental data and determines optimal courses of action based on pre-defined rules, optimization criteria, or machine learning models. This may include tasks such as route planning, obstacle avoidance, resource optimization, and time-sensitive responses in dynamic environments. The component also translates decisions into executable instructions for downstream systems, ensuring alignment between intent and action.

Actuators

Actuators represent the execution layer of the autonomous system, responsible for carrying out physical actions in response to decisions generated by the system. These components convert digital commands into mechanical or operational outputs. The selection and configuration of actuators are context-dependent and influenced by factors such as required force, speed, precision, and environmental conditions.

Learning module

The learning module is a critical enabler of adaptability and continuous improvement within autonomous systems. It aggregates data from sensors, historical performance records, and external data sources (including cloud-based systems), and applies analytical and machine learning techniques to identify patterns, optimize performance, and refine internal models.

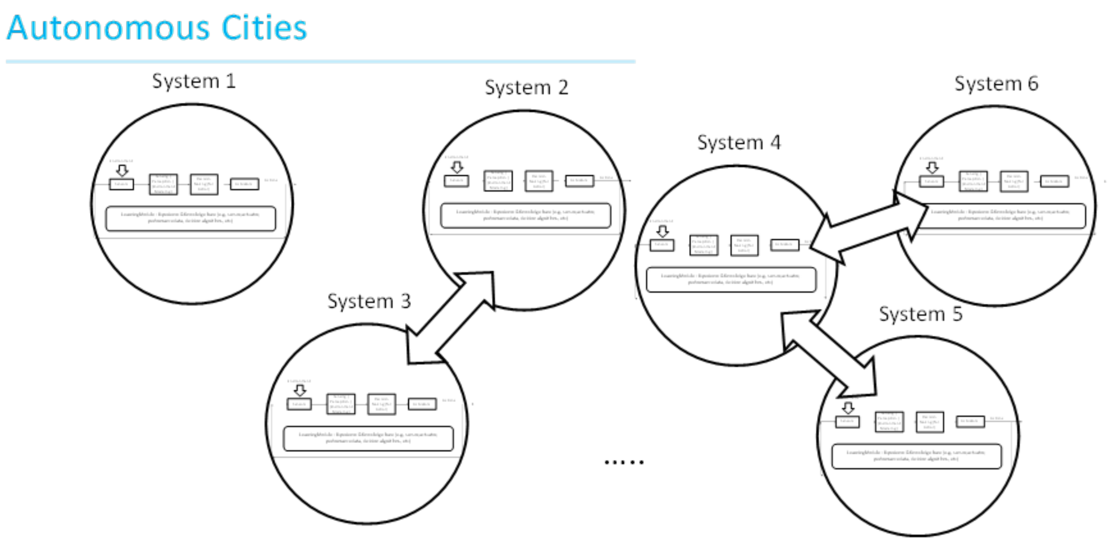
Over time, this module enhances key system capabilities such as perception accuracy and decision-making efficiency by continuously updating models based on new data and operational feedback. It also facilitates the accumulation of contextual knowledge specific to the system's operating environment, including recurring patterns (e.g., traffic flows), environmental conditions, and optimal control strategies. As such, the learning module plays a central role in advancing the intelligence, robustness, and adaptability of autonomous systems.

Autonomous systems within urban contexts

It is important to note that, in principle, each system, and sub-system, within a city can be conceptualized and implemented as an autonomous system. In this regard, the systems illustrated in Figure 1 may be mapped onto autonomous system architectures, as depicted in Figure 2. Figure 3 further extends this conceptualization by illustrating a city in which all constituent systems operate as autonomous systems.

Accordingly, an autonomous city may be defined as a city in which its constituent systems are, to varying extents, autonomous in their operation and decision-making capabilities.

Figure 3 – City composed of autonomous systems (Source: Author)



Assessment of autonomy in urban systems

The degree of autonomy across city systems is not uniform and may vary significantly depending on technological maturity, operational requirements, regulatory constraints, and risk considerations. This variability necessitates the development of structured approaches to assess and evaluate the extent of autonomy within individual systems and across the city as a whole.

To this end, an assessment framework may be established to quantify the level of autonomy exhibited by a given system. Such a framework can incorporate standardized criteria and metrics to evaluate capabilities across sensing, decision-making, execution, and learning dimensions.

Given that a city comprises a collection of interconnected systems, the overall level of city-wide autonomy may be derived through aggregation methodologies. These may include, for example, the calculation of an arithmetic average or a weighted average of autonomy levels across systems, with weights reflecting system criticality, scale or impact.

For example, a city may assign higher weights to systems such as energy, mobility, public safety and water because failures in these systems may have broad consequences. Conversely, lower-risk administrative systems may be assigned lower weights. The weighting methodology should be transparent and should be reviewed periodically as city priorities and system maturity evolve.

- An autonomy score may be assigned to an individual system.
- An aggregate score may be assigned to a cluster of systems such as infrastructure, mobility, public services or safety.
- A city-level autonomy score may be calculated across selected systems using a transparent aggregation methodology.
- Scores should be interpreted alongside risk, readiness and governance maturity; a high autonomy score is not automatically desirable if safeguards are weak.

Such scoring mechanisms enable benchmarking, comparative analysis, and informed decision-making, supporting policymakers and city administrators in strategically advancing the transition towards autonomous urban systems.

4 Autonomous cities assessment framework

The assessment framework provides a structured method to evaluate the extent and quality of autonomy in city systems. It is designed to support planning, benchmarking, prioritization and governance. The framework intentionally combines technical capability assessment with broader public-sector considerations, recognizing that public trust and institutional accountability are as important as technological performance.

In this section, first an assessment framework will be developed and suggested for autonomous systems. Then the same concept will be extended to cities which are essentially interconnected systems.

4.1 Autonomous systems assessment framework

The assessment framework suggested in this section adopts the high-level model of autonomous systems depicted in Figure 2. The extent of autonomy is then applied to each of the components in Figure 2; namely sensing (or perception), decision-making, and learning (or adapting). This is referred to as component-specific aspects. Additionally, the assessment framework also entails general aspects (which are not component-specific) such as self-sufficiency and adherence to AI principles.

It is important to note that some of the components in Figure 2 may not exist in every system (e.g., a system may operate or make decisions in a pre-defined manner regardless of changes in its environment and does not have learning capability from previous experience). Alternatively, a system can sense its environment and based on those values can make decisions through pre-defined algorithms or heuristics; however, it does not learn or modify its decisions based on experience, rather based on only sensing its environment during operation).

In this assessment framework, it is assumed that the extent of autonomy of each of the components of a system during operation can be measured as one of the four distinct levels:[1]

- **Human-in-the-Loop (HITL) / Human-centric**
    - The Paradigm: The AI system lacks execution authority. It acts strictly as an advisor, recommender or data processor. A human must actively review the AI's output, make the final decision and initiate the action.
    - Governance Implication: The operational risk is inherently mitigated by direct human oversight. The liability and final agency rest entirely with the human operator.
- **Human-on-the-Loop (HOTL) / Supervised autonomy**
    - The Paradigm: The AI system has the authority to execute tasks and make operational decisions autonomously within tightly defined parameters. However, a human operator continuously monitors the process.
    - Governance implication: This is the standard transitionary phase for deploying physical or embodied AI in dynamic, public environments. It requires highly resilient digital infrastructure and low-latency monitoring capabilities to ensure the human can intervene in time.
- **Human-in-Command (HIC)**
    - The Paradigm: Frequently utilized in policy and legal frameworks, this level focuses on macro-level strategic oversight rather than micro-level tactical intervention. The human

[1] The assumption of four distinct levels is made for simplification. The practitioners of this assessment framework can enhance and expand on the proposed simple method to include more granular and higher number of levels (including a continuum in some cases).

dictates the rules of engagement, algorithmic constraints, and overall objectives, but is completely removed from the loop of individual actions.

- Governance implication: Crucial for establishing the legal boundaries and compliance requirements before deployment, ensuring the system operates entirely within a pre-approved ethical and operational envelope.

- **Human-out-of-the-Loop (HOOTL) / Fully autonomous**
  - The Paradigm: The AI system functions entirely independently. It perceives the environment, formulates policy or operational decisions, and executes actions without any human intervention, monitoring, or ability to override during the operational cycle.
  - Governance Implication: This level requires the highest degree of rigorous pre-deployment testing, strict data sovereignty and residency compliance, and comprehensive legislation to address liability, auditing, and systemic resilience.

A. Component-specific assessment – Environment perception

It aims to understand the system's ability to sense and interpret its surroundings.

Key issues to inquire are:

- Does the system have sensors (physical or digital) to gather information about its environment?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.
- Does the system integrate data from multiple sensors to create a comprehensive picture of the environment?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.
- Does the system process and understand the environment information to make informed decisions?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.
- If the system is connected to other system(s)[2], does it gather and process inputs from other system(s)?

[2] City systems tend to be interconnected to each other and they have ripple effects as discussed earlier.

- No
- Yes – the extent of autonomy with respect to four levels can be assessed.

B. Component-specific assessment – Decision-making

It aims to understand the system's ability to define its goals and make decisions.

Key issues to inquire are:

- Does the system dynamically modify and adapt its objectives (performance goals)?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.
- Does the system make choices and plans to achieve its objectives?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.
- Does the system consider different options and potential consequences before acting?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.

C. Component-specific assessment – Learning

It aims to understand the system's ability to learn and adapt from past experience and behaviour.

Key issues to inquire are:

- Does the system adjust its operation or decision-making based on past interactions with the environment?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.
- Does the system improve its performance over time through some form of learning?
  - No
  - Yes – the extent of autonomy with respect to four levels can be assessed.

D. General assessment – AI principles adherence

It aims to understand the system's adherence to AI principles during design and operation.[3] Please note that the principles included in this section are excerpted from the U4SSC deliverable "Guiding principles for artificial intelligence in cities".

Key considerations are:

Principle	Assessment question
Lawful	Does the system adhere to all relevant laws, regulations, policies, procurement requirements and sector-specific obligations?
Privacy preserving	Does the system respect and protect privacy during design, development, deployment and operation?
Fair and inclusive	Is the system designed to avoid bias and discrimination and to incorporate the needs of diverse communities?
Explainable and transparent	Can system behaviour, decisions and limitations be explained to operators, oversight bodies and affected persons where appropriate? Are these transparently available?
Accountable	Are responsibilities clearly assigned across owners, operators, suppliers, decision-makers and oversight bodies?
Safe and secure	Does the system operate safely and securely under normal, abnormal and adversarial conditions?
High-performing and robust	Does the system achieve target performance under varying inputs, conditions and operational scenarios?
Assessed for impact and sustainability	Have social, economic, environmental and institutional impacts been assessed?
Enabling human autonomy	Does the system preserve human agency and allow meaningful intervention, review or appeal where appropriate?

E. General assessment – Additional considerations

It aims to understand various aspects of the system from a more general perspective.[4]

Key considerations are:

Consideration	Assessment focus
Risk	Level of harm or disruption that may arise from system failure, misuse, bias or unintended consequences.
Readiness	Availability of infrastructure, skills, procedures, budgets, governance structures and public communication mechanisms.

[3] The practitioners can grant scores to reflect adherence to AI principles (different levels of scoring can be applied depending on the intended precision and granularity during the assessment).

[4] The practitioners can grant scores to various indicated considerations (different levels of scoring can be applied depending on the intended precision and granularity during the assessment).

Consideration	Assessment focus
People-centricity	Alignment of system objectives with resident needs, service quality, accessibility and lived experience.
Accessibility	Ability of all users, including persons with disabilities, to access and benefit from the system.
Interoperability	Ability to connect with other systems securely, reliably and meaningfully.
Human intervention	Frequency, clarity and effectiveness of human monitoring, override and escalation.
Accuracy and effectiveness	Rate at which the system performs correctly and achieves intended outcomes.
Adaptability	Speed and reliability with which the system responds to new data or changing conditions.
Ease of use	Usability for operators, administrators and affected users.
Reliability	Ability to perform consistently without errors, failures or unacceptable downtime.
Maintainability	Ease of diagnosing, repairing, updating and improving the system over time.

4.2 Autonomous cities assessment

City administrators should begin by defining the systems and subsystems to be included within the scope. The scope may initially focus on a limited set of high-priority systems such as mobility, energy or public services, and may expand over time. For each selected system, the autonomous systems assessment framework should be applied consistently.

The results may then be aggregated to produce system-cluster and city-level insights. Aggregation should not obscure material risks. A city may achieve a relatively high average score while still having weak governance in a critical domain. For this reason, aggregate autonomy scores should be accompanied by risk profiles, readiness assessments and qualitative findings.

5 Autonomous cities assessment framework methodology

The methodology is designed to translate assessment into action. It guides city administrators through a practical sequence that begins with baselining and ends with evaluation and learning. The approach can be applied to individual systems, system clusters or the city as a whole.

The four steps are as follows:

I. Assess the current extent of autonomy in the city's systems (Baselining).

II. Determine the target autonomous state for city's systems.

III. Implement actions to enhance the extent of autonomy in city's systems.

IV. Evaluate Results.

The following briefly explains the four-step methodology.

I. Assess the current extent of autonomy in the city's systems (Baselining)

The baseline assessment establishes the current state of autonomy for selected systems. It should identify existing capabilities, data sources, operating procedures, governance arrangements, risks, dependencies and gaps. The baseline should be evidence-based and should involve system owners, technical experts, policy teams, risk and compliance functions, service operators and, where appropriate, external stakeholders.

- Define the scope of systems and subsystems to be assessed.
- Map system objectives, owners, users, data flows and dependencies.
- Assess component-specific autonomy across perception, decision-making and learning.
- Assess general requirements, including legal compliance, safety, privacy, accessibility, interoperability and maintainability.
- Document evidence, assumptions, uncertainties and known limitations.
- Identify immediate risks requiring mitigation before further autonomy is introduced.

Two representative tables incorporating the assessment factors introduced in Section 4 are shown in the Appendix. It can be used by city administrators to assess its current status, or baseline, with respect to those factors for each and every system included in the scope.[5]

II. Determine the target autonomous state for city's systems

The target state defines the desired level of autonomy for each system or subsystem. It should be based on public value, risk appetite, technical feasibility, resource availability, institutional readiness and stakeholder needs. The target should not assume that the highest degree of autonomy is always preferable. In some systems, human-augmented decision-making may be more appropriate than fully autonomous operation.

- Identify the public outcomes the system is expected to improve.
- Determine the appropriate autonomy level for each function.
- Determine the targets with respect to component-specific and general assessment aspects.
- Define performance, safety, privacy, inclusion and accountability requirements.

[5] The contents of the tables in the Appendix are high-level and, in some cases, may require further clarification and description by the city administrators to determine its current status.

- Prioritize systems and gaps based on impact, feasibility and risk.
- Establish measurable targets and timelines.
- Confirm governance and funding requirements for implementation.

III. Implement actions to enhance the extent of autonomy in city's systems

Implementation should be organized around a structured action plan. The plan should identify interventions, responsible owners, timelines, resources, dependencies, procurement needs and performance metrics. Actions may include data integration, sensor deployment, AI model development, interoperability upgrades, cybersecurity improvements, regulatory changes, operating procedure redesign, workforce training and public communication.

A phased approach is recommended. Pilot deployments can be used to test performance, validate risks and refine governance arrangements before scaling. Critical systems should require more rigorous testing, assurance and contingency planning than low-risk systems.

Action Area	Examples of implementation measures
Data and sensing	Improve data quality, deploy sensors, standardize metadata, establish data-sharing protocols.
Technology and integration	Upgrade APIs, integrate systems, validate AI models, strengthen cybersecurity and resilience.
Governance	Clarify ownership, define escalation paths, establish audit requirements and approve operating boundaries.
People and skills	Train operators, build AI assurance capability, update procurement and contract management practices.
Regulation and policy	Define permissible autonomy levels, liability arrangements, privacy requirements and safety obligations.
Engagement	Communicate system purpose, safeguards and benefits; gather feedback from users and affected communities.

IV. Evaluate results

Evaluation determines whether the actions implemented have achieved the intended improvements and whether risks remain acceptable. Evaluation should combine quantitative performance measurement with qualitative feedback and independent review where appropriate.

- Measure performance against baseline and target indicators.
- Assess whether autonomy improvements have delivered public value.
- Review safety, privacy, cybersecurity, inclusion and accessibility outcomes.
- Collect feedback from operators, users, affected communities and system owners.
- Conduct cost-benefit and risk-benefit analysis.

- Identify unintended consequences and required corrective actions.
- Capture lessons learned and update the autonomy roadmap.

The methodology should be applied iteratively. As systems mature and new capabilities emerge, cities should reassess autonomy levels, update risk controls and revise target states. Iteration also allows cities to expand scope gradually, moving from pilot systems to broader urban operating models while maintaining governance discipline.

6 Conclusion

Autonomous systems are increasingly relevant to the future of smart sustainable cities. They can improve urban performance by enabling systems to sense, decide, act and learn with increasing sophistication. However, their value depends on responsible governance, not on autonomy alone.

This guide has advanced three central arguments. First, cities are complex systems of systems, and autonomy must be assessed in the light of interdependencies and ripple effects. Second, autonomous cities should be understood as a continuum of capabilities across systems rather than a single technology-driven end-state. Third, the responsible advancement of urban autonomy requires structured assessment, phased implementation, continuous evaluation and strong safeguards.

There is no one-size-fits-all model for autonomous cities. Each city must define its own priorities, governance arrangements, risk thresholds and implementation pathway. The proposed framework can support this process by providing a common language, assessment structure and methodology for moving from aspiration to practical action.

Cities should exercise caution in transitioning towards high levels of autonomy, particularly in systems that affect safety, rights, essential services or vulnerable populations. Human oversight, accountability, transparency, cybersecurity, accessibility and sustainability should remain integral throughout the lifecycle. Where these conditions are met, autonomous systems can become an important enabler of more responsive, resilient, inclusive and sustainable urban governance.

- Autonomy should be pursued only where it advances clear public value.
- Human oversight should remain proportionate to risk and impact.
- City-level autonomy requires assessment of systems, subsystems and interdependencies.
- Governance, readiness and public trust are as important as technical capability.
- Assessment should be iterative and should inform investment, regulation and operational improvement.

Appendix

This appendix includes two tables capturing the factors developed in Section 4 for assessing autonomous systems.

Table A.1: Component-specific assessment factors

Assessment Factor	0 Not Present	1 Human-in-the-Loop (HITL) / Human-Centric	2 Human-on-the-Loop (HOTL) / Supervised Autonomy	3 Human -in - Command (HIC)	4 Human-out-of-the-Loop (HOOTL) / Fully Autonomous	Evidence / Comments
Does the system have sensors, physical or digital, to gather information about its environment?						
Does the system integrate data from multiple sensors or sources to create a comprehensive picture of the environment?						
Does the system process and interpret environmental information to make informed decisions?						
If connected to other systems, does it gather and process inputs from those systems?						

Assessment Factor	0 Not Present	1 Human-in-the-Loop (HITL) / Human-Centric	2 Human-on-the-Loop (HOTL) / Supervised Autonomy	3 Human -in - Command (HIC)	4 Human-out-of-the-Loop (HOOTL) / Fully Autonomous	Evidence / Comments
Does the system dynamically modify and adapt its objectives or performance goals?						
Does the system make choices and plans to achieve its objectives?						
Does the system consider different options and potential consequences before acting?						
Does the system adjust its operation or decision-making based on past interactions with the environment?						
Does the system improve its performance over time through learning or feedback?						

Table A.2: General assessment factors

General Assessment Factor	Score	Evidence / Comments
Lawful: The system adheres to relevant laws, regulations, policies and sector-specific requirements.		
Privacy preserving: The system protects privacy during design, deployment and operation.		
Fair and inclusive: The system avoids discrimination and accounts for diverse needs.		
Explainable and transparent: System behaviour and decisions can be explained and transparently available where appropriate.		
Accountable: Responsibilities are assigned across design, deployment and operation.		
Safe and secure: The system operates safely and is protected against cybersecurity threats.		
High-performing and robust: The system maintains target performance across expected conditions.		
Assessed for impact and sustainability: Social, economic and environmental impacts have been assessed.		
Enabling human autonomy: Human agency, review and intervention are preserved where needed.		
Risk: The level of risk posed by the system has been identified and managed.		
Readiness: Infrastructure, skills, processes and institutional capacity are sufficient.		
People-centricity: Objectives and requirements align with resident and user needs.		
Accessibility: All users, including persons with disabilities, can access and use the system.		
Interoperability: The system integrates securely and reliably with other systems.		
Human intervention: Intervention frequency and escalation rules are appropriate.		
Accuracy and effectiveness: The system performs correctly and achieves intended outcomes.		
Adaptability: The system responds appropriately to new data and changing conditions.		
Ease of use: The system is usable by operators and relevant stakeholders.		
Reliability: The system performs consistently without unacceptable errors or downtime.		
Maintainability: The system can be diagnosed, repaired, updated and improved over time.		

Bibliography

Axmann, B., & Harmoko, H. (2020). Robotic process automation: An overview and comparison to other technology in Industry 4.0. In *2020 10th International Conference on Advanced Computer Information Technologies (ACIT)* (pp. 559–562). IEEE. https://doi.org/10.1109/ACIT49673.2020.9208907

Azar, A. T., & Koubaa, A. (2023). Artificial intelligence for robotics and autonomous systems applications. Springer.

Batty, M. (2013). *The new science of cities*. MIT Press. https://doi.org/10.7551/mitpress/9399.001.0001

Faisal, A. I. M., Yigitcanlar, T., Kamruzzaman, L., & Currie, G. (2019). Understanding autonomous vehicles: A systematic literature review on capability, impact, planning and policy. *Journal of Transport and Land Use, 12*. https://doi.org/10.5198/jtlu.2019.1405

Ribeiro, J., Lima, R., Eckhardt, T., & Paiva, S. (2021). Robotic process automation and artificial intelligence in Industry 4.0: A literature review. *Procedia Computer Science, 181*, 51–58. https://doi.org/10.1016/j.procs.2021.01.104

Vicente, D. M., Pereira, R. S., & Leal, A. A. (2024). Legal aspects of autonomous systems: A comparative approach. *ICASL: International Conference on Autonomous Systems and the Law Conference Proceedings*.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., & Wen, J.-R. (2024). A survey on large language model based autonomous agents. *Frontiers of Computer Science, 18*, Article 186345. https://doi.org/10.1007/s11704-024-40231-1

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

ISBN 978-92-61-43001-6

Published in Switzerland
Geneva, 2026

Photo credits: @AdobeStock