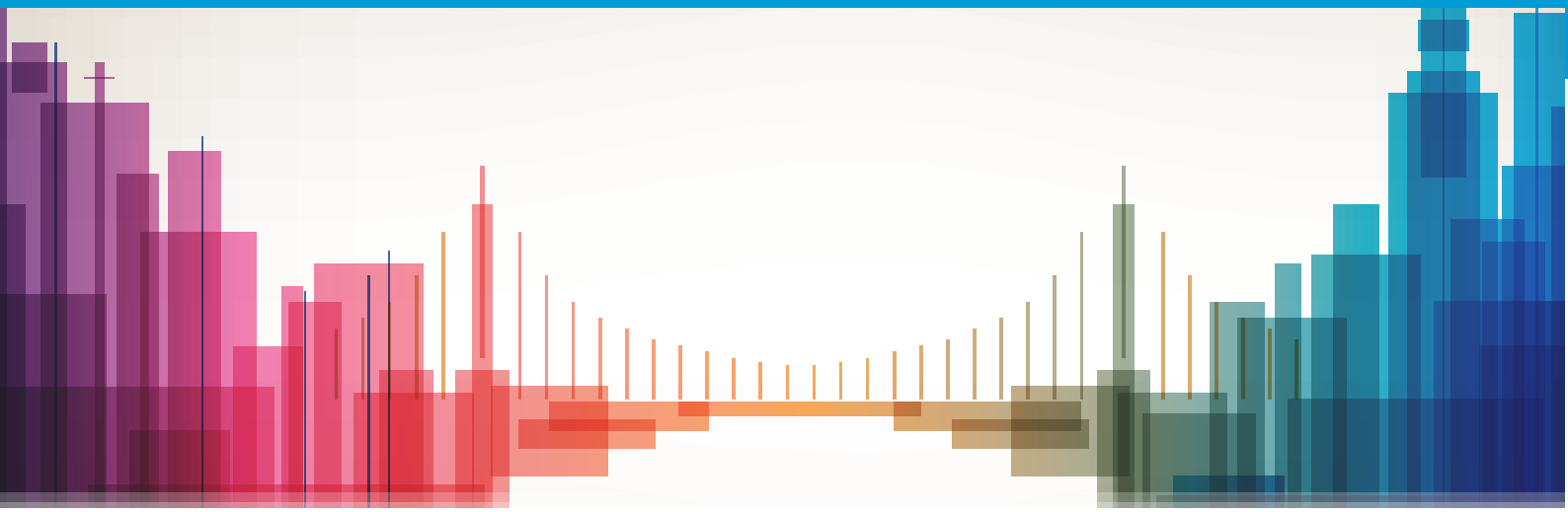
Good practices for city AI agents

A framework for responsible and people-centric deployment of AI agents in cities



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

UNEP FINANCE 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
AGENT
AI
Generate

Good practices for city AI agents

A framework for responsible and people-centric deployment of AI agents in cities

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: Adam Obeid (Independent Researcher), Anna Badimo (Sigmateq), Eduardo Alencar (Brazilian Chamber of Deputies), Grace Jimenez (Researcher), Jose Manuel Pérez Marzabal (Member of the Barcelona Bar Association), Kwi-Gon Kim (Seoul National University), Paul Wilson (Smart Cities World), Rajagopal Tampi (AIPathfinder), Shalini S Gopalkrishnan (San Francisco Bay University), Silvia Blanco (SB Analytics) and Val Wise (Citiverse Council).

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 extend their 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

978-92-61-43831-9 (electronic version)
978-92-61-43841-8 (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.


GuestGrid.js
in your browser.

Contents

Abbreviations and acronyms viii

Executive Summary ix

1 Purpose, Audience and Scope 1
- 1.1 Purpose 1
- 1.2 Audience 1
- 1.3 Scope 1
- 1.4 Recommended policy status 1

2 AI Agents in Government and Cities 2
- 2.1 Working definition 2
- 2.2 How AI agents differ from traditional automation 2
- 2.3 Illustrative city use cases 3
- 2.4 Agent lifecycle 4

3 Guiding Principles 5

4 AI Agent Classification and Risk Tiering 6
- 4.1 Autonomy levels 6
- 4.2 Risk tiers 6
- 4.3 Factors that increase risk 7
- 4.4 Uses requiring potential prohibition or exceptional approval 7

5 Recommended Good Practices for AI Agents 8
- 5.1 Public purpose and policy alignment 9
- 5.2 Scope, mandate and authorized use 9
- 5.3 Governance, ownership and accountability 10
- 5.4 Human oversight and decision authority 11
- 5.5 Legal, ethical and rights-based safeguards 12
- 5.6 Data governance, privacy and confidentiality 13
- 5.7 Security, access control and cyber-resilience 13
- 5.8 System integration and tool-use controls 14
- 5.9 Interoperability and cross-government coordination 15
- 5.10 Testing, validation and assurance 15
- 5.11 Procurement and vendor management 16
- 5.12 Transparency, public communication and trust 17
- 5.13 Public AI literacy and digital inclusion 18
- 5.14 Operational monitoring, auditability and incident management 18
- 5.15 Workforce readiness and change management 19
- 5.16 Performance and public value measurement 20
- 5.17 Lifecycle Management and continuous improvement 20

5.18 Intellectual property, training-data provenance and content licensing 21
6 Pathway to Responsible Autonomous City Capabilities 23
Annex A: Use Case Intake and Approval Template 27
Annex B: Risk Classification Matrix 28
Annex C: Pre-Deployment Readiness Checklist 29

List of figures

Figures

Figure 1: A U4SSC Framework for Responsible and People-centric City AI Agents x
Figure 2: Traditional Automation vs AI Agents 2
Figure 3: Illustrative City Use Cases 3
Figure 4: Recommended Good Practices for AI Agents 8
Figure 5: Pathway to Responsible Autonomous City Capabilities 23

Abbreviations and acronyms

Abbreviation	Full form
AI	Artificial intelligence
API	Application programming interface
ICT	Information and communication technology
IGO	Intergovernmental organization
IP	Intellectual property
ISBN	International Standard Book Number
KPI	Key performance indicator
OECD	Organisation for Economic Co-operation and Development
RPA	Robotic process automation
U4SSC	United for Smart Sustainable Cities

Executive Summary

Cities are entering a new era of digital public service. Artificial intelligence (AI) agents can do more than answer questions: within defined limits, they can organize tasks, use approved tools and help deliver services. For mayors, the opportunity is practical: faster responses to residents, simpler administrative procedures, better coordination between departments and more responsive urban services. The test of success is not how advanced the technology is, but whether it improves people's lives.

Consider the possibilities: an agent could help residents navigate municipal services, flag missing documents in permit applications, support maintenance teams or provide timely information during transport disruptions. These are illustrative uses, not promises of automatic results. Cities should start with a clearly defined public problem, assess whether an AI agent is the right solution and measure outcomes such as service quality, accessibility, staff time and public satisfaction.

Greater capability also brings greater responsibility. An agent may access sensitive information, connect government systems or influence decisions affecting rights, safety and essential services. Public authorities must remain accountable. Residents should know when AI is involved, be able to reach a person and have meaningful ways to question or challenge consequential outcomes. Human oversight must be real: officials need the information, authority and time to intervene, override or stop an agent.

This United for Smart Sustainable Cities (U4SSC) framework for responsible and people-centric city AI agents gives mayors and city administrations a practical approach to responsible deployment. It distinguishes four risk tiers: low, moderate, high and critical with safeguards proportionate to potential harm. Higher-risk uses demand stronger assessment, testing, auditability and human control; critical uses require exceptional authorization and assurance. The framework also sets out 18 areas of good practice, from public purpose and legal authority to procurement, cybersecurity, inclusion and continuous monitoring.

The immediate leadership agenda is clear: choose a public-service problem worth solving; identify a named accountable owner; define what the agent may and may not do; classify its risk; require independent scrutiny where appropriate; and establish monitoring, incident response and a clear route for residents to obtain human assistance. The annexes provide a use case intake template, risk classification matrix and pre-deployment checklist to help turn these commitments into operational decisions.

The message for city leaders is simple: more autonomy requires more accountability. Deploy AI agents because they deliver demonstrable public value not simply because the technology makes them possible.

Figure 1: A U4SSC Framework for Responsible and People-centric City AI Agents



The headline: "Because we can" is not a reason to deploy an AI agent. "Because it serves the public, and someone is accountable" is.

This is the mayor's rulebook for AI agents: software that does not just answer questions but plans, uses tools and takes action. It gives city leaders a practical test for what to approve, what to supervise and what to refuse. The core idea is simple: the more an agent can affect people's rights, safety or essential services, the tougher the rules.

What mayors should know:

- **Match the rules to the risk.** Four tiers run from low (internal drafting) to critical (emergency response, core infrastructure, legal entitlements). Critical agents need executive sign-off, fail-safes and independent review.
- **Every agent needs a named owner.** Business, technical, data and risk owners, all listed in a citywide agent registry. If nobody is accountable, do not launch.
- **A human signature is not human oversight.** Reviewers need the time, information and authority to say no. Rubber-stamping an automated decision does not count.
- **Residents have rights at the counter.** They should know when they are dealing with AI, get an explanation, reach a human, and be able to appeal.

- **Don't get locked in or left exposed.** Contracts should guarantee audit rights, exit options and protection against copyright claims over training data.
- **Autonomy is earned in stages.** Start with agents that assist staff, move to supervised workflows, and allow conditional autonomy only in tightly bounded areas such as traffic or utilities.

The mayor's move: Ask one question of every proposed agent: can the city reconstruct what happened if it gets something wrong? Then adopt the paper's ready-made tools: an intake template, a risk matrix and a pre-launch checklist.

The soundbite: More autonomy demands more governance, not less.

City leader imperative

A city AI agent should not be deployed merely because it is technically feasible. It should be deployed when it serves a clear public purpose, operates within an authorized mandate, protects rights and interests, and can be monitored, corrected and a designated individual can be held accountable.

Leadership priority	What city leaders should require
Public value	A clear statement of the service, policy or operational outcome the AI agent is intended to improve.
Authority	A defined mandate specifying what the AI agent may recommend, decide or execute.
Accountability	Named business, technical, data and risk owners for every deployed AI agent.
Human oversight	Documented decision points for AI agent where human review, approval, override or escalation is required, if any.
Safety and legality	Risk assessment, privacy review, security testing and compliance validation of AI agent before launch.
Auditability	Logs of AI agent instructions, data access, tool calls, outputs, user interactions, escalations and incidents.
Continuous review	Performance monitoring, incident learning, revalidation and retirement criteria of AI agent.



1 Purpose, Audience and Scope

This section defines the intent of the framework and the city context in which it should be applied.

1.1 Purpose

This document sets out good practices for deploying, overseeing and continuously improving AI agents in government, with a focus on cities and urban services. It helps city leaders turn general responsible-AI principles into practical requirements for AI agents.

1.2 Audience

- Mayors, city managers, governors and elected or appointed urban leaders responsible for public-service outcomes.
- Chief digital, data, AI, innovation, information, security, risk and technology officers in city public institutions.
- Public-service, infrastructure, mobility, planning, emergency management, utilities and administrative leaders considering agentic AI use cases.
- Regulators, auditors, legal officers, procurement authorities and ethics bodies responsible for oversight and assurance.
- Programme managers and delivery teams responsible for deploying, supervising or procuring government AI agents.

1.3 Scope

- Applies to AI agents developed, procured, operated, or authorized by public institutions in cities, including those delivered by third parties, on-premises or in the cloud.
- Covers internal AI agents, public-facing AI agents, workflow AI agents, decision-support AI agents, tool-using AI agents and multi-agent AI systems, among others.
- Addresses governance, ethics, risk management, data use, security, integration, assurance, oversight, monitoring and lifecycle management.
- Does not replace applicable law, sector regulation, procurement rules, professional standards, cybersecurity requirements or institutional mandates of public institutions in cities.

1.4 Recommended policy status

Cities may adopt this framework as voluntary guidance, a mandatory internal policy, a procurement requirement or a regulatory reference. A city that makes it binding should state clearly the legal authority, the accountable institution, the scope, the compliance process and the review cycle.

2 AI Agents in Government and Cities

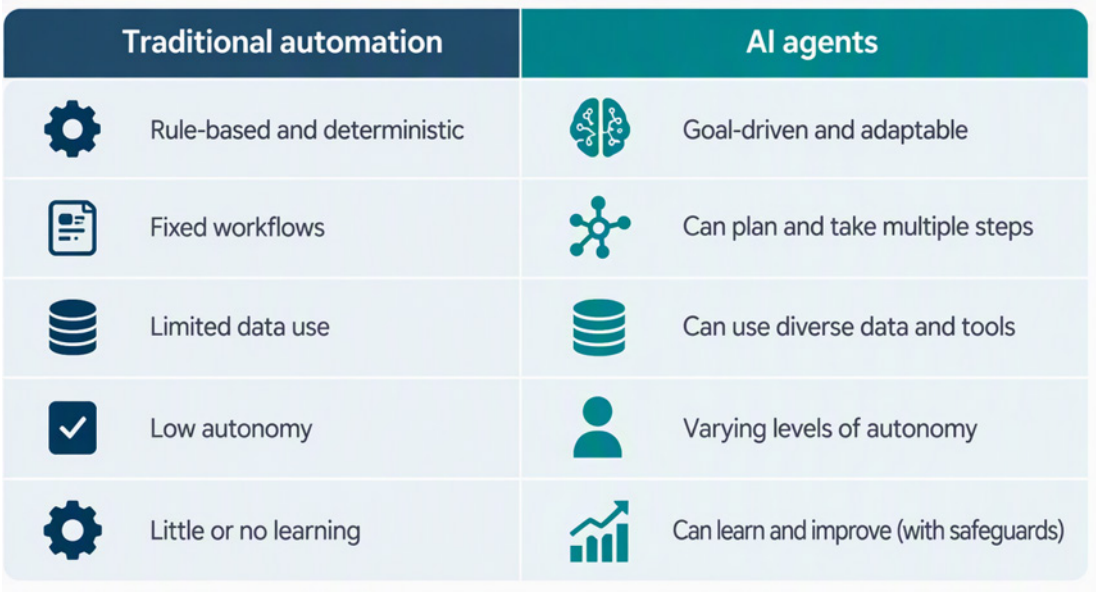
This section explains AI agents in plain language and distinguishes them from traditional automation.

2.1 Working definition

City AI agent: AI-enabled software, deployed by or for a public institution in a city, that can interpret information, pursue a goal it has been given, plan a sequence of tasks, communicate with users or systems, use approved tools, and recommend or carry out actions within authorized limits.

2.2 How AI agents differ from traditional automation

Figure 2: Traditional Automation vs AI Agents



Dimension	Traditional automation	AI agent
Operating logic	Follows predefined rules and workflows.	Can plan, adapt and sequence actions based on goals, context and constraints.
Interaction model	Typically form-based or system-to-system.	May converse with users, call tools, coordinate tasks and request clarification.
Flexibility	Limited to known inputs and pathways.	Can handle more varied inputs, while requiring stronger safeguards.
Risk profile	Risks are generally tied to workflow design.	Risks include data misuse, unsafe tool use, hallucinated outputs, prompt injection, unapproved autonomy and accountability gaps.

Dimension	Traditional automation	AI agent
Governance need	Change control and process controls.	Risk tiering, human oversight, AI agents' registry, tool permissions, audit logs and continuous monitoring.

2.3 Illustrative city use cases

Figure 3: Illustrative City Use Cases



Use case type	Possible AI agent role	Typical governance concern
Public information and service navigation	Answer questions, guide residents to services, summarize requirements and route requests. Collect minimum information required for the task.	Accuracy, language accessibility, culture, privacy, disclosure of AI use and escalation to human support.
Permitting and licensing	Pre-screen applications, identify missing documents, draft recommendations and coordinate follow-ups.	Due process, human review, recordkeeping, explainability and non-discrimination.
Urban operations	Monitor events, recommend maintenance actions and coordinate field-service workflows.	System reliability, safety, data quality and operational continuity.
Mobility and transport	Support disruption response, information dissemination, scheduling or optimization recommendations.	Public safety, transparency, real-time monitoring and access to authorized documents, and sources contingency planning.
Emergency and resilience management	Assist with situational summaries, resource coordination and scenario analysis.	Human command authority, accuracy, timeliness and fail-safe controls.
Administrative productivity	Draft documents, summarize cases, prepare meeting notes and support procurement or finance workflows.	Confidentiality, data leakage, records management and approval controls.

Important distinction

An AI agent can be useful without being "completely" autonomous. Many high-value government AI agents can remain assistive or advisory, especially where legal rights, entitlements, safety or essential services are involved.

2.4 Agent lifecycle

Lifecycle stage	Key decision	Minimum output
Identify	Is there a legitimate public purpose?	Use case definition, goal and a brief problem statement.
Classify	What is the risk tier and autonomy level?	Risk classification and oversight requirements. Ethical and legal considerations.
Design	What data, tools, prompts and controls are needed?	Design specification and control plan.
Validate	Is the agent safe, accurate, secure and compliant enough to pilot?	Testing and assurance report.
Deploy	Can it operate within approved limits?	Launch approval, monitoring plan and incident pathway.
Operate	Is it delivering value without unacceptable harm?	Monitoring dashboard, audit logs and review records.
Improve or retire	Should the agent be scaled, changed, restricted or withdrawn?	Revalidation, change approval or retirement decision.

3 Guiding Principles

These principles are the common foundation for public trust, responsible innovation and effective governance of AI agents in cities.

Principle	Meaning for government AI agents
Public value and human dignity	AI agents should support legitimate public objectives and achieve favorable civic outcomes (e.g., improve outcomes for people, communities and the environment).
Legality and mandate	AI agents should operate only within the legal authority and institutional mandate of the deploying public body.
Human accountability	A public institution and named officials remain accountable for AI agent design, development, deployment, use, impacts and correction.
Proportionality and necessity	The level of autonomy, data access and intervention of the AI agents should be proportionate to the problem being addressed. Least privilege principle should be applied (e.g., for access, autonomy, decision-making).
Safety, reliability and fallback	AI agents should be tested, monitored and constrained, in order to reduce foreseeable risks to people, infrastructure and services. Real-time systems should incorporate a fallback mechanism to ensure continuity of operation in the case of catastrophic failure.
Transparency and contestability	People should receive appropriate notice, explanation and channels for review or challenge when AI agents affect them. Incident reporting portals should exist using standard formats (e.g., OECD common reporting framework for AI incidents).
Privacy and data stewardship	AI agents should use data lawfully, minimally, securely and for defined purposes. Least-privilege principle should be applied (e.g., for access, autonomy, decision-making).
Fairness and inclusion	AI agents should be designed to avoid discrimination, exclusion and unequal access to public services.
Security and resilience	AI agents should be protected against misuse, cyberthreats, prompt injection, unauthorized actions and service disruption.
Interoperability and openness	AI agents should support standards-based integration, portability, auditability and avoidance of unnecessary lock-in.
Sustainability and continuous improvement	AI agents should be reviewed over time for performance, risk, social impact, environmental cost and ongoing public value.

Principles in practice

For city leaders, the key question is not solely whether an AI agent is innovative, but whether it is duly authorized, necessary, safe, accountable, inclusive, and capable of delivering clear and demonstrable civic outcomes and public benefits.

4 AI Agent Classification and Risk Tiering

A risk-based approach lets cities encourage innovation while keeping the strongest safeguards for the agents with the greatest impact.

4.1 Autonomy levels

Level	Description	Illustrative control requirement
Level 0 - Basic support	AI agent retrieves or organizes information but does not generate recommendations or actions.	Standard data, security and records controls.
Level 1 - Assistive agent	AI agent drafts, summarizes, translates or helps a human complete a task.	Human review before external use where accuracy matters.
Level 2 - Advisory agent	AI agent recommends options or decisions but does not execute them.	Documented human decision-maker and explanation requirements.
Level 3 - Supervised execution	AI agent can execute approved actions after human confirmation or within narrow rules.	Approval gates, logs, rollback and escalation path.
Level 4 - Conditional autonomy	AI agent can execute defined actions without case-by-case approval in a constrained domain.	Enhanced monitoring, independent assurance and emergency stop controls.
Level 5 - Broad autonomy	AI agent can pursue objectives across domains with limited direct human intervention.	Requires special legal authority, exceptional safeguards with meticulous oversight.

4.2 Risk tiers

Risk tier	Typical characteristics	Governance implication
Tier 1 - Low	Internal productivity support; no direct effect on rights, benefits, enforcement, safety or essential services.	Lightweight approval, standard controls and periodic review.
Tier 2 - Moderate	Public-facing support or workflow assistance; limited personal data; outputs inform but do not determine service outcomes.	Use case approval, documented oversight, accuracy testing and user escalation.
Tier 3 - High	Influences materially important decisions, handles sensitive data, interacts with critical workflows or may affect access to services.	Formal risk assessment, legal and privacy review, human-in-the-loop controls, audit logging and independent validation.

Risk tier	Typical characteristics	Governance implication
Tier 4 - Critical	Impacts public safety, emergency response, essential infrastructure, liberty, legal entitlements or large-scale public resource allocation.	Executive authorization, exceptional assurance, continuous monitoring, fail-safe mechanisms and external or independent review.
Excluded or prohibited	Uses that are unlawful, incompatible with human rights and human safety, lack a clear mandate, or cannot be made safe or accountable.	Do not deploy. Escalate duly to competent legal, ethics and oversight authorities. Can be deployed once all mitigation measures have been fully implemented, thereby reducing the risk to an acceptable level.

4.3 Factors that increase risk

- The AI agent affects eligibility, entitlements, enforcement, inspections, penalties, housing, assets, mobility, health, education, employment, taxation or access to essential services.
- The AI agent accesses sensitive personal data, confidential government information, security information or large integrated datasets.
- The AI agent can execute actions through APIs, robotic process automation bridges, databases, payment systems, identity platforms, infrastructure systems or communication channels.
- The AI agent is public facing, used at scale, deployed across multiple agencies or likely to affect vulnerable or marginalized groups.
- The AI agent relies on third-party models, opaque vendor systems, external data sources or systems that cannot be adequately audited.
- Errors could create safety risks, reputational harm, public confusion, service denial, discriminatory outcomes or legal exposure.
- The city has limited ability to access, control, observe or maintain the AI agent.

4.4 Uses requiring potential prohibition or exceptional approval

- AI agents used for purposes that are unlawful under applicable law or outside the mandate of the deploying city institution.
- AI agents that make binding legal determinations about individuals without meaningful human review, clear legal authority and appeal mechanisms.
- AI agents that conduct indiscriminate surveillance, social scoring or profiling in ways that undermine rights, dignity or equal treatment.
- AI agents that access or manipulate critical systems without secure authorization, auditability, operational controls and emergency stop mechanisms.
- AI agents used to deceive people about the nature of public authority, impersonate officials without disclosure, or suppress access to public services.

5 Recommended Good Practices for AI Agents

Figure 4: Recommended Good Practices for AI Agents



AI agents can greatly strengthen cities' ability to deliver services, coordinate across agencies and respond to new public needs. That potential is realized only when agents are designed, run and governed in ways that protect human safety and the public interest, ensure accountability and uphold fundamental rights. Without clear guidelines, agents can introduce bias, opacity, unauthorized decisions, misuse of data or system-wide failures that undermine public trust and the integrity of institutions.

This section sets out good practices for designing, deploying and governing AI agents in cities. They guide city leaders, public servants, technical teams and oversight bodies through the whole life of an agent, from planning and procurement to monitoring, evaluation and continuous improvement. Each practice keeps autonomous or semi-autonomous agents aligned with human safety, public purpose, legal authority, ethical standards and human oversight.

The guidelines cover eighteen core areas. Together they give cities an integrated framework for deploying AI agents responsibly, safely and effectively, while keeping accountability and public confidence.

5.1 Public purpose and policy alignment

Objective: Ensure that every AI agent is tied to a legitimate public purpose, service objective or operational need while ensuring human safety.

Recommended practices:

- Define the public problem, including any human-safety aspect, the people affected and the expected outcome before choosing a technical solution.
- Document the policy, service or operational objective that the agent supports.
- Show that an AI agent is necessary, or clearly better than simpler digital tools.
- Check whether the AI agent could unintentionally shift public priorities, reduce access or create hidden burdens for residents.
- Align AI agent deployment with broader city strategies for digital government, inclusion, resilience, fallback processes, sustainability and service quality.

Question for city leaders

What public value will this AI agent create, and how will the city know whether that value has been achieved?

5.2 Scope, mandate and authorized use

Objective: Define the operational boundaries within which the AI agent may act.

Recommended practices:

- Specify authorized users, operating environments, service channels, data sources, tools, permitted actions and prohibited actions.

- Separate what the agent may do on its own, what a human can override, what needs human approval and what it may only recommend.
- Translate policy boundaries into technical controls such as tool permissions, access scopes, approval gates and rate limits.
- Maintain a current agent card or profile describing the agent purpose, limitations, risk tier, owners and escalation contacts.

Question for city leaders

Would a member of the public, an auditor and an operator understand what this agent is and is not allowed to do?

5.3 Governance, ownership and accountability

Objective: Assign clear responsibility for each AI agent throughout its lifecycle.

Recommended practices:

- Designate a business owner accountable for the public-service outcome and acceptable use of the agent.
- Designate a technical owner responsible for architecture, model configuration, tool integration and operational performance.
- Designate data owners responsible for data access, quality, sharing, retention and lawful use.
- Designate a risk or assurance owner(s) responsible for risk classification, review, escalation and independent challenge.
- Establish a cross-functional review and oversight committee for moderate, high and critical risk agents.
- Maintain a public institution-wide and/or city-wide AI agents registry covering all deployed and pilot agents.
- Create an agent incident reporting portal in a standardized format (e.g., OECD AI incident reporting framework).

Question for city leaders

Who is accountable when an AI agent performs incorrectly, causes harm, or performs outside its intended boundaries?

5.4 Human oversight and decision authority

Objective: Ensure meaningful human control over decisions and actions that affect rights, safety, essential services or public trust.

Recommended practices:

- Define whether the AI agent is assistive, advisory, supervised, conditionally autonomous or broadly autonomous.
- Require human approval for high-stakes binding decisions, high-impact recommendations, external communications and irreversible actions, unless clearly authorized otherwise.
- Ensure that human reviewers have sufficient information, time, authority and training to judge the quality of the AI agent's outputs.
- Provide ways to override, pause, roll back and stop the agent in an emergency, and to give redress. Operational AI agents also need a fallback: rollback handles a single failed process, but a fallback system keeps the service running if the whole system fails.
- Avoid superficial human review where staff simply approves AI agent outputs without understanding or scrutiny.

The table below lists possible models of human oversight.

Model	Description	Appropriate use examples
Human-in-the-loop	Humans approve each material action before execution.	High-risk, irreversible, rights-affecting, financial, legal, or external-action workflows.
Human-on-the-loop	Human supervises ongoing operation through alerts and dashboards and can intervene.	Moderate-risk tasks, reversible tasks, bounded customer-service workflows.
Human-over-the-loop	Humans set policy and reviews samples, metrics, and incidents periodically.	Low-risk, internal, reversible, non-sensitive automated tasks.
Human-out-of-the-loop	No active human oversight during operation.	Minimal-risk and tightly constrained tasks.

Question for city leaders

Is the human role meaningful, or is the city simply adding a human signature to an automated decision?

5.5 Legal, ethical and rights-based safeguards

Objective: Protect legality, due process, fairness and the rights and interests of residents, visitors and businesses.

Recommended practices:

- Map AI agent use to applicable laws, public powers, records requirements, procurement rules and sector-specific obligations.
- Identify whether the AI agent may affect safety, rights, benefits, obligations, enforcement, inspection, allocation of resources or access to public services.
- Provide notice, explanation, review and appeal pathways where AI agents influence important outcomes.
- Assess foreseeable impacts on vulnerable groups, minorities, people with disabilities and digitally excluded populations.
- Require ethical review for high-impact or socially sensitive use cases, including proportionality and alternatives analysis.
- Where an AI agent makes, or largely determines, a decision with legal or similarly significant effects on a person, four safeguards apply whatever the agent's formal autonomy level: notice, meaningful human intervention, the chance to express a point of view and the ability to contest the decision. Where several agents work together, apply these safeguards to the outcome of the whole process, not only to individual agents.
- Do not rely on consent as the legal basis for AI agents built into essential or practically unavoidable public services, where consent cannot be freely given or realistically withdrawn. Rely on a clear statutory or public-task basis. Balance it with strict purpose limitation, a meaningful opt-out or non-digital alternative where feasible, and effective redress.
- Treat data-protection, fundamental-rights and conformity assessments as living documents tied to operations, not one-off paperwork before launch. Repeat them when the model is updated, new data sources are added, the scope expands or performance drifts.

Question for city leaders

Could a person understand, challenge and obtain remedy for an important outcome influenced by this agent?

5.6 Data governance, privacy and confidentiality

Objective: Ensure that AI agents use data lawfully, securely, accurately and only for defined public purposes.

Recommended practices:

- Apply data minimization: provide the AI agent only the data needed for its approved task.
- Confirm lawful basis, purpose limitation, retention, sharing and disclosure requirements before deployment.
- Validate data quality, provenance, completeness, timeliness and bias risks for datasets used by the AI agent.
- Separate training, testing and operational data environments where appropriate.
- Apply strict and legally mandated controls for sensitive personal data, confidential government information, protected records and operational security data used by the AI agent.
- Prevent data leaking out through prompts, tool outputs, logs, integrations or third-party systems.

Question for city leaders

What data can the AI agent access, why is that access necessary, and who can verify that it is not being misused?

5.7 Security, access control and cyber-resilience

Objective: Protect AI agents and connected systems against misuse, compromise, manipulation and unauthorized action.

Recommended practices:

- Give every AI agent, tool, API, database, service account and human operator only the access its role requires (least privilege).
- Use strong identity, authentication, authorization, secrets management and credential rotation.
- Test for prompt injection, malicious file content, data poisoning, unauthorized tool calls and privilege escalation.
- Separate AI agent permissions by environment, use case and risk tier; avoid broad reusable credentials.

- Log and monitor events, including security-relevant ones, such as failed tool calls, unusual data access, abnormal task volume and policy violations.
- Develop continuity plans for AI agent outage, compromise, vendor failure or degraded model performance.

Question for city leaders

If the AI agent were manipulated, what systems could it reach and how quickly could the city contain the incident?

5.8 System integration and tool-use controls

Objective: Ensure that AI agents interact safely with government systems, digital platforms and operational tools.

Recommended practices:

- Use approved API wrappers, tool registries and integration gateways instead of direct unmanaged access to back-end systems.
- Define tool schemas, allowed parameters, validation rules, transaction limits and conditions for human approval.
- Use sandbox environments for testing tool calls before production deployment.
- Implement transactional safeguards such as confirmations, duplicate detection, rollback, idempotency and audit trails.
- Treat robotic process automation bridges as a temporary fix for legacy systems. Monitor them closely, as they are inherently fragile and may fail when underlying applications, such as screen layouts or user interfaces, change.
- Keep data synchronized so that AI agents work from current, consistent and authoritative information.

Question for city leaders

Can the city prove what tools the AI agent used, what actions it took, and whether each action was authorized?

5.9 Interoperability and cross-government coordination

Objective: Enable AI agents to work across agencies and service domains without creating fragmented or duplicative governance.

Recommended practices:

- Adopt common data standards, metadata practices, identity services, APIs and event-sharing mechanisms where feasible.
- Define roles for shared platforms, sector agencies, local authorities and central digital or data authorities.
- Create interoperability requirements for agents that operate across permits, mobility, planning, utilities, public safety or social services.
- Use shared registries, semantic layers, service catalogues and consent or authorization mechanisms where applicable.
- Coordinate incident reporting, audit requirements and escalation pathways across participating entities.

Question for city leaders

Will this AI agent strengthen whole-of-government service delivery, or will it create another isolated digital silo?

5.10 Testing, validation and assurance

Objective: Confirm before launch that the AI agents are sufficiently accurate, safe, secure, fair and fit for purpose.

Recommended practices:

- Test with realistic scenarios, edge cases, multilingual interactions, vulnerable-user scenarios and adversarial prompts.
- Measure accuracy, task completion, refusal quality, escalation behaviour, bias, latency and user experience.
- Validate outputs against authoritative sources, especially for legal, financial, safety or service eligibility information.
- Conduct security testing and red-teaming proportionate to risk tier and tool access.

- Require sign-off from business, technical, data, legal, privacy, security and risk functions before production release for high-tier risk agents.
- Document known limitations, residual risks, mitigations and conditions for deployment.

Question for city leaders

What evidence shows that the AI agent is ready for public or operational use, and who has independently challenged that evidence?

5.11 Procurement and vendor management

Objective: Ensure that third-party AI agents and platforms meet the city's public-sector standards for trust, transparency, security and continuity.

Recommended practices:

- Include requirements for data residency, data use, confidentiality, audit rights, security controls, incident notification and exit arrangements.
- Require vendor disclosure of model capabilities, limitations, dependencies, subcontractors, update practices and service-level commitments.
- Assess whether the city can monitor performance, export records, change providers and maintain continuity if the vendor relationship ends.
- Avoid procurement arrangements that prevent meaningful audit, portability, explainability or tracing accountability.
- Require contractual controls for model updates, prompt changes, tool permissions, training on government data and reuse of city data.
- Require intellectual-property warranties and indemnities covering claims arising from the provider's training data and from generated outputs, backed by adequate financial standing or insurance (see 5.18).
- Require the right to audit, or to obtain independent third-party attestation of, training and fine-tuning processes, as far as legitimate trade secrets allow.
- Treat portability and exit as legal-risk controls. Require open, standards-based interfaces and export formats, so that the city can replace a model or provider without rebuilding the service and can keep enforcing its ethical and compliance standards.
- When commissioning or co-developing agent components, make a deliberate, documented licensing choice (e.g., copyleft, permissive or public-domain models). Weigh the public interest

in keeping publicly funded improvements public against the need to work with the private sector.

Question for city leaders

Would the city still be able to explain, audit and operate the public service if the vendor changed its model, terms or availability?

5.12 Transparency, public communication and trust

Objective: Promote public understanding of when, why and how government AI agents are used.

Recommended practices:

- Inform users when they are interacting with an AI agent or when an AI agent materially supports a service outcome.
- Provide accessible explanations of the AI agent purpose, limitations, data use, human review options and complaint pathways.
- Publish a public-facing register or summary for significant AI agents, subject to security and confidentiality considerations.
- Use plain language, inclusive design and multiple communication channels to reach diverse communities.
- Monitor public feedback, complaints and trust indicators as part of operational governance.
- Extend public registers beyond the systems the city itself develops or buys. Significant agents run by other public bodies or private actors in the urban space, under city authorization, should also be visible.
- Move from describing transparency to practising it. Add to each register entry the agent's purpose, legal basis, risk tier and human-oversight model and, for higher-risk agents, aggregate performance and incident indicators. Keep entries current, so that the published map matches the systems actually deployed.
- Link every register entry to channels people can use: understandable explanations, complaint and appeal routes, and a contact point for the accountable owner.

Question for city leaders

Would residents see the AI agent as a tool for better public service, or as an opaque barrier between them and the city government?

5.13 Public AI literacy and digital inclusion

Objective: Ensure city residents and businesses (customers in general) understand how to interact with AI agents, recognize their capabilities and limitations, and use AI-enabled public services safely, confidently, and equitably.

Recommended practices:

- Develop public education initiatives that explain what city AI agents are, how they are used, and what they can and cannot do.
- Provide accessible guidance on interacting with AI agents, including how to verify information, recognize AI-generated responses, and escalate to a human when needed.
- Ensure AI literacy resources are inclusive, multilingual, and accessible to people with disabilities, older adults, digitally excluded populations, and other underserved communities.
- Clearly communicate residents' rights when interacting with AI agents, including transparency, privacy protections, opportunities for human review, and complaint or appeal mechanisms.
- Monitor levels of public understanding, trust, and confidence in AI-enabled city services, and use feedback to improve communication, education, and service design.

Question for city leaders

Will residents have the knowledge and confidence to use this AI agent effectively, understand its limitations, and know when and how to seek human assistance?

5.14 Operational monitoring, auditability and incident management

Objective: Maintain effective oversight once AI agents are deployed in real services or operations.

Recommended practices:

- Maintain logs for prompts, instructions, data access, outputs, tool calls, decisions, approvals, overrides, errors and incidents.

- Monitor key performance, fairness, safety, security, service-quality and cybersecurity indicators in real time or at defined intervals.
- Define thresholds that trigger human review, automated suspension, rollback, fallback or escalation.
- Define incident categories, response times, reporting obligations, root-cause analysis and corrective actions, together with cybersecurity monitoring and management.
- Review audit logs periodically to detect drift, misuse, unexpected behaviour or emerging risks.

Question for city leaders

Can the city reconstruct what happened if the agent produces an incorrect, harmful or contested outcome?

5.15 Workforce readiness and change management

Objective: Prepare public servants to use, supervise and challenge AI agents responsibly.

Recommended practices:

- Define new roles, responsibilities and standard operating procedures for human-agent collaboration.
- Train staff on AI agent capabilities, limitations, escalation, override, privacy, cybersecurity and records obligations.
- Ensure staff understand that AI outputs require professional judgment, especially in high-impact services.
- Engage employee representatives and operational teams early when AI agents affect work design, workload or accountability.
- Use pilots and feedback loops to improve usability and reduce operational burden.

Question for city leaders

Are city public servants equipped to govern the AI agent, or are they expected to trust a system they do not understand?

5.16 Performance and public value measurement

Objective: Measure whether AI agents improve public outcomes without creating unacceptable costs or harm.

Recommended practices:

- Define baseline indicators before AI agent deployment, including service time, error rates, backlog, user satisfaction, cost and accessibility.
- Track both productivity benefits and public-value or civic outcomes of AI agent, not only technical usage metrics.
- Measure adverse outcomes such as complaints, escalations, service denial, unequal impact, rework and incidents related to AI agent.
- Compare AI agent-assisted processes and their cost against human-only or conventional automation alternatives where feasible.
- Report AI agent performance to leadership and oversight bodies at a frequency proportionate to risk and scale.
- Set standard key performance indicators (KPIs) so that AI agents are evaluated consistently across city agencies. They should cover, among others: operational efficiency (e.g., response time, task completion, cost savings); service quality (e.g., accuracy, user satisfaction, first-contact resolution); governance and safety (e.g., hallucination rate, human escalation rate, policy violations, incidents); fairness and inclusion (e.g., accessibility, equitable outcomes across demographic groups); public value (e.g., improved service accessibility, trust and resident outcomes); and availability (e.g., downtime for technical or cybersecurity reasons).

Question for city leaders

Is the AI agent improving the service in ways that matter to the public, or only increasing internal automation metrics?

5.17 Lifecycle Management and continuous improvement

Objective: Ensure AI agents remain safe, effective and aligned with public purpose over time.

Recommended practices:

- Define what triggers revalidation: model updates, data changes, new integrations, expanded scope, incidents or a material drift in performance.

- Review AI agent risk classification periodically and when the operating context changes.
- Maintain version control for prompts, policies, model settings, tool permissions and evaluation results.
- Define retirement criteria for AI agents that no longer deliver value, cannot be governed effectively or create unacceptable risk.
- Capture lessons from incidents, audits, user feedback and staff experience to improve future deployments.

Question for city leaders

What would cause the city to restrict, redesign or retire an AI agent?

5.18 Intellectual property, training-data provenance and content licensing

Objective: Ensure that the models, knowledge sources and generated outputs underpinning city AI agents rest on a lawful and demonstrable provenance basis, and that the city does not assume undisclosed intellectual-property, data-protection or liability exposure created upstream in the AI supply chain.

Recommended practices:

- Apply this practice area in proportion to risk. For low-risk internal agents, a provider's attestation of lawful training may be enough. For public-facing or high-risk agents, require documentary evidence, output testing and contractual remedies.
- Require providers of the underlying models to attest that training and fine-tuning complied with copyright and text-and-data-mining rules, and to provide evidence for higher-risk agents. This includes respecting rights-holders' machine-readable opt-outs or reservations where the law provides for them.
- Obtain meaningful documentation of where training content came from, in proportion to the agent's risk tier. Some jurisdictions already require training-content summaries from providers of general-purpose models.
- Where the city fine-tunes, grounds or augments a model with municipal data, confirm a lawful basis for any personal data involved, apply minimization and purpose limitation, and assess and document how data-protection roles are split between the city and the provider.
- Before public release, assess the risk that outputs infringe rights or allow individuals to be re-identified. Use proportionate similarity checks, filters or citations for public-facing generated content. Treat any output that allows a person to be re-identified as personal data, with full safeguards.

- Allocate intellectual-property risk by contract. Require warranties and indemnities covering claims arising from training data and from the model's characteristic generation patterns, backed by adequate financial standing or insurance. Clarify who owns, and who may reuse, the prompts, fine-tuned models, embeddings and outputs produced with city data.
- Where fit for purpose, prefer models and knowledge sources with clean or open provenance (openly licensed corpora, public-sector information, city-owned content). They reduce legal exposure and support local knowledge ecosystems.
- Keep records that let the city show afterwards which model version, knowledge sources and licences lay behind a given public output or decision, in line with the auditability practices in 5.14.

Question for city leaders

If a rights-holder, resident or auditor challenged either the provenance of this AI agent's training data or the originality of its outputs, could the city answer with evidence rather than with its vendor's assurances?

6 Pathway to Responsible Autonomous City Capabilities

Figure 5: Pathway to Responsible Autonomous City Capabilities



AI agents offer cities a structured path from digitally enabled systems to urban operations that are more coordinated, more anticipatory and responsibly autonomous. This does not replace human governance. It deliberately strengthens public capacity. With strong governance, robust data infrastructure and transparent oversight, AI agents help cities deliver services more efficiently, spot emerging risks earlier, coordinate better across agencies and respond faster to residents' needs.

The path should be phased and iterative. Each stage builds on the institutional, technical and ethical foundations of the one before. Autonomy is not an end in itself. The goal is measurable public value: greater resilience, inclusiveness, sustainability, accountability and trust in public institutions.

Stage 1: Assisted public operations

At first, AI agents act mainly as intelligent assistants to public servants. They retrieve information, draft documents, coordinate administration and carry out preliminary analysis. In this phase:

- AI agents support human decisions; they do not make decisions themselves.
- Human operators retain full authority over all operational outcomes.
- Governance focuses on clear use case definitions, data quality standards, and pilot-level assurance mechanisms.
- Training programmes begin building staff capacity to work alongside AI agents and tools effectively.

The focus is on building confidence, showing early public value and putting basic safeguards in place without adding significant operational risk.

Stage 2: Supervised workflow execution

As governance matures, data quality improves and systems connect better, cities can move to supervised workflows. AI agents begin to support or partly automate routine processes such

as permit triage, service allocation, evidence-based routing and information sharing between agencies. In this stage:

- AI agents operate under defined rules, human review points, and escalation protocols.
- Human oversight remains central. Automated steps need confirmation or approval where risk or impact is significant.
- Assurance mechanisms expand to include performance monitoring, bias detection, and audit trails.
- Interoperability standards enable more seamless coordination across departments and jurisdictions.

This stage needs better data synchronization, secure integration and semantic layers, and institutional mandates that make responsibilities, accountability and the limits of automated action clear.

Stage 3: Conditional autonomy in selected functions

Once digital infrastructure is mature, data governance is robust, oversight is reliable and public trust is established, cities may allow conditional autonomy in selected functions. In carefully defined areas such as traffic management, utility optimization, environmental monitoring or emergency coordination, AI agents may act with limited autonomy under explicit constraints:

- Autonomy is conditional: bounded by policy rules, legal authority and safety thresholds.
- Human oversight remains present through monitoring, fallback mechanisms, and periodic review.
- Systems are transparent and auditable, with clear logs of decisions, actions and exceptions.
- Public accountability is maintained through institutional responsibility, redress mechanisms, and community engagement.

Conditional autonomy is never absolute. Governance, safety and accountability requirements always constrain it.

Foundational conditions for autonomous capabilities

Autonomous city capabilities must rest on six critical conditions:

- **Mature digital infrastructure:** Reliable, secure, and scalable infrastructure capable of supporting real-time data exchange, integration and resilience.
- **High-quality, synchronized data:** Consistent, accurate, and interoperable data across agencies, with clear ownership, stewardship, and protection standards.

- **Secure integration layers:** Protected APIs, communication protocols, and middleware that ensure safe, auditable, and controlled data flow.
- **Clear institutional mandates:** Defined roles, responsibilities, access rules, and authority for AI deployment, operation, and oversight.
- **Accountable oversight mechanisms:** Independent audit, monitoring, and review processes, including human-in-the-loop safeguards where risk or impact is significant.
- **Human review and redress:** Meaningful mechanisms for residents and stakeholders to challenge decisions, seek corrective action, and obtain explanations.

Principles guiding the pathway

Cities should follow a set of principles for how autonomous capabilities are developed and deployed:

- **Public value first:** Autonomy is pursued only when it delivers measurable benefits in service delivery, efficiency, resilience, or equity.
- **Human-centred governance:** AI augments human decision-making; it never replaces prudent oversight or institutional accountability.
- **Transparency and explainability:** Systems should be understandable, auditable, and capable of providing reasoned explanations for decisions.
- **Safety and risk mitigation:** Robust safeguards prevent harm, manage failures, and ensure reliable fallback mechanisms.
- **Inclusiveness and equity:** AI agents should be designed to advance fairness, reduce disparities, and serve all residents equitably.
- **Trust and accountability:** Public trust is maintained through clear ownership, oversight, enforcement, and redress mechanisms.

A deliberate and measured approach

Progress toward autonomous city capabilities should be deliberate, evidence-based and adaptive. Cities should not deploy faster than their governance, infrastructure or oversight can support. Each stage should be assessed against clear performance criteria, public-value metrics and risk assessments.

AI agents are a strategic means to better governance, not a technological imperative. Autonomous capabilities should emerge only where the city has built the necessary foundations and shown that responsible deployment will strengthen public trust, resilience and service quality.

By following this pathway, cities can realize the transformative potential of AI while safeguarding good governance, human rights and public accountability, and keep the future of urban operations anchored in the public interest.

Leadership caution

Autonomous city capabilities necessitate more rigorous, not diminished, standards of governance. As AI agents gain greater operational independence, the responsibilities of city authorities correspondingly intensify, particularly in ensuring robust accountability, stringent safety measures, transparency in decision-making, and the preservation of meaningful human oversight and control.

Annex A: Use Case Intake and Approval Template

Cities may adapt the following template for initial screening and approval of AI agent initiatives.

Field	Information to capture
Use case title	Clear name of the proposed AI agent or agentic workflow.
Sponsoring entity	Department, agency or public body responsible for the use case.
Public purpose	Service, policy or operational objective the agent is intended to support.
Affected users	Residents, businesses, visitors, public servants, suppliers or other stakeholder groups affected.
Agent function	Assistive, advisory, supervised execution, conditional autonomy or other role.
Data sources	Datasets, systems, documents and real-time feeds the agent will access.
Tools and integrations	APIs, databases, RPA bridges, workflow tools, communications channels or operational systems.
Potential impact	Possible effects on rights, benefits, obligations, safety, essential services or public resources.
Risk tier	Preliminary classification and rationale.
Human oversight	Review, approval, override and escalation arrangements.
Assurance evidence	Testing, security review, privacy assessment, legal review and user acceptance evidence.
Launch decision	Approved, approved with conditions, deferred, rejected or escalated.

Annex B: Risk Classification Matrix

This matrix helps cities assign a preliminary risk tier. Adapt it to local law and to social and institutional risk appetite.

Risk factor	Lower-risk indicator	Higher-risk indicator
Impact on people	No material effect on rights, benefits, obligations or access to services.	Material effect on rights, eligibility, enforcement, penalties, safety or essential services.
Data sensitivity	Public or low-sensitivity internal data.	Sensitive personal data, confidential records, security data or integrated cross-agency datasets.
Autonomy level	Drafting, summarization or internal support only.	Tool execution, automated workflow decisions or conditional autonomy.
Scale	Small internal pilot with limited users.	Public-facing, city-wide or multi-agency deployment.
Transparency	Users can easily identify and correct errors.	Errors may be difficult to identify and remediate due to the limited explainability of reasoning agents. Affected people may not know the agent influenced an outcome.
Reversibility	Errors are easy to detect and correct.	Errors are difficult to reverse or may cause legal, financial, safety or reputational harm.
System connectivity and access	No direct access to operational systems.	Access to APIs, registries, payment systems, identity systems, infrastructure or communication channels.
Vulnerable groups	No foreseeable disproportionate impact.	Possible impact on children, elderly people, people with disabilities, migrants, low-income groups or digitally excluded communities.
Technology architecture	Little or no impact of technology architecture on inferencing	Higher impact by the underlying technology architectures, which may influence and propagate erroneous inferences produced by the agent or model.
Real-time or non-real time deployments	Low risk for non-real time systems	Higher risk due to ultra-low latencies for response
Content and training provenance	Model and knowledge sources have documented openly licensed or city-owned provenance; outputs are internal drafts subject to human review.	Opaque third-party model with undisclosed training corpus; agent publishes generated content directly to the public or produces outputs derived from sensitive or protected records.

Annex C: Pre-Deployment Readiness Checklist

This checklist should be completed before a pilot or production launch.

Area	Readiness question	Status
Public purpose	Is the agent linked to a clear service, policy or operational objective?	Yes / No / Partial
Mandate	Is the deploying institution authorized to operate the agent for this purpose?	Yes / No / Partial
Risk tier	Has the agent been classified by risk tier and autonomy level?	Yes / No / Partial
Ownership	Are business, technical, data and risk owners assigned?	Yes / No / Partial
Human oversight	Are approval, override, escalation and suspension procedures defined?	Yes / No / Partial
Data governance	Has lawful basis, minimization, quality, retention and confidentiality been reviewed?	Yes / No / Partial
Security	Have access controls, secrets, prompt-injection risks and tool permissions been tested?	Yes / No / Partial
Tool use	Are APIs, RPA bridges and workflow actions constrained and logged?	Yes / No / Partial
Testing	Has scenario, edge-case, bias, accuracy, user and adversarial testing been completed?	Yes / No / Partial
Transparency	Are user notice, explanation, complaint and appeal channels ready where needed?	Yes / No / Partial
Operations	Is monitoring, incident response, rollback and support coverage in place?	Yes / No / Partial
IP and provenance	Have training data provenance, licensing, output-infringement risk and contractual indemnities been reviewed and documented?	Yes / No / Partial
Launch approval	Has the appropriate authority approved deployment and any conditions?	Yes / No / Partial



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

ISBN 978-92-61-43831-9

Published in Switzerland
Geneva, 2026

Photo credits: @AdobeStock