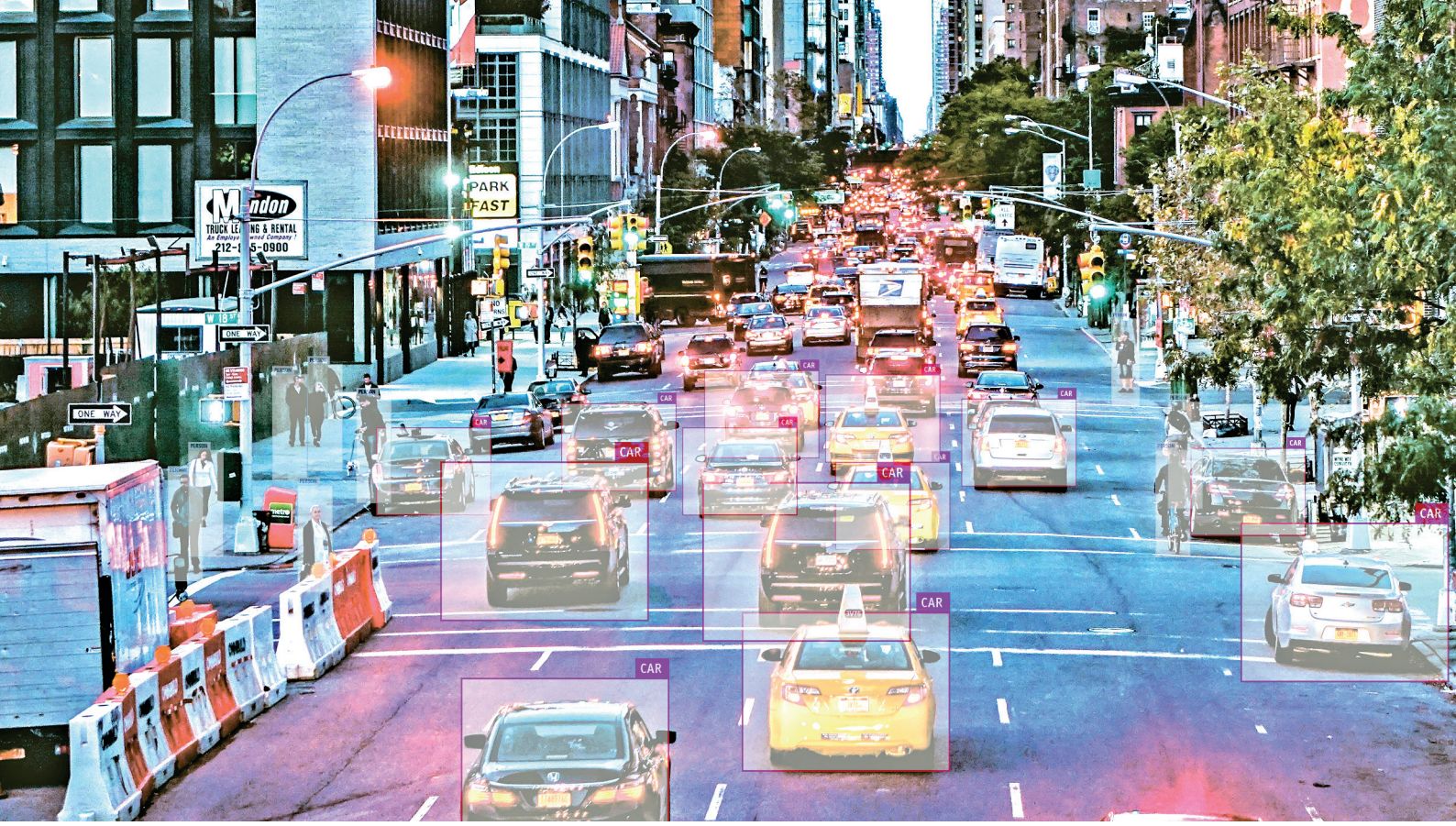

Ethical urban AI in practice

Policy mechanisms to establish local governance frameworks

Marta Galceran-Vercher and Alexandra Vidal D'oleo (eds.)

CIDOB GLOBALCITIES
BARCELONA CENTRE FOR INTERNATIONAL AFFAIRS

Ajuntament de Barcelona

CIDOB MONOGRAPHS 89

Ethical urban AI in practice

Policy mechanisms to establish local governance frameworks

Marta Galceran-Vercher and Alexandra Vidal D'oleo (eds.)

CIDOB
BARCELONA CENTRE FOR INTERNATIONAL AFFAIRS

GLOBALCITIES

Ajuntament de Barcelona

CIDOB MONOGRAPHS 89

CIDOB edicions
Elisabets, 12
08001 Barcelona
Tel.: 933 026 495
www.cidob.org
cidob@cidob.org

Print: QP Print Global Services
ISBN: 978-84-18977-28-2
Legal Deposit: B 1974-2025

Barcelona, December 2024

Cover photo:
Vehicles on pathway surrounded by high-rise buildings. 10th Ave.
https://unsplash.com/photos/vehicles-on-pathway-surrounded-by-high-rise-buildings-3UnFTjhsWWA

CONTENTS

ABOUT THE AUTHORS

Tanya Álvarez

Is a researcher for the Mobile World Capital Foundation's Observatory. She leads research on digital inclusion and the use of automated decision-making systems in the public sector. She advocates for an interdisciplinary perspective of how technology impacts society. She has a degree in Art History from Swarthmore College and a Master's degree in Cultural Heritage Management from the University of Barcelona.

Josuan Eguiluz Castañeira

Holds a degree in Law and ICT from the University of Deusto and a Dual Master's in Access to the Legal Profession and Intellectual Property, New Technologies and Data Protection from Esade. He began his professional career as a lawyer in the Intellectual Property, New Technologies and Data Protection Department at Cuatrecasas. Currently, he serves as Legal Counsel at Adevinta. He is also a lecturer in digital law at Esade, Deusto and The Legal School, among others. He is pursuing an Industrial PhD in AI, High-Risk Areas and Fundamental Rights at Universitat Pompeu Fabra, Barcelona Supercomputing Center and Infojobs/Adevinta.

Carlos Fernández Hernández

Holds a degree in law and is currently a PhD candidate at Universidad Carlos III de Madrid. After practising as a lawyer for a number of years, he has devoted his professional career to the publication of legal content for professionals in the sector, always under the "La Ley" brand. Since 2013, he has specialised in the field of law and technology, initially at the legal news portal *Noticias Jurídicas* and, since 2016, at *Diario La Ley Ciberderecho*, where he has published over 1,500 articles, reports and interviews. Carlos has also been the editor of the journals *Derecho Digital e Innovación* and *La Ley Privacidad*. He has delivered specialised lectures on these subjects at various universities and postgraduate centres and contributed to several collaborative books.

Marta Galceran-Vercher

Is a Senior Research Fellow with the Global Cities Programme at CIDOB (Barcelona Centre for International Affairs), and a part-time lecturer in international relations at Pompeu Fabra University and CEI-International Affairs. She is a political scientist, holds a PhD in International Relations

from UPF and a master's (awarded with special distinction) from the University of Warwick. Her research focuses on city diplomacy, urban alliances and global governance, particularly in the field of digital transitions and technological humanism. She has authored several policy papers, academic articles and other publications on these topics. She also leads the research at the Global Observatory of Urban Artificial Intelligence (GOUAI), within the Cities Coalition for Digital Rights. Before joining CIDOB, she held the position of Program Coordinator of the Smart City Expo World Congress and Senior Consultant at Anteverti, where she advised local governments and international organisations (i.e. UN-Habitat, IADB, the European Commission) on internationalisation and urban innovation policies.

Shazade Jameson

Is a Senior Digital Governance Consultant for public interest organisations. She specialises in urban public administrations. As a social science researcher, she shifts perspectives on digital policy and builds bridges for effective implementation. She was lead co-author of the UN-Habitat and Mila Quebec AI Institute report "AI and Cities: Risks, Applications and Governance". She is also a PhD candidate analysing technology policy in Singapore, as part of the Global Data Justice project funded by the European Research Council at the Tilburg Institute of Law, Technology, and Society. She has collaborated with CIDOB on policy research for the Global Observatory of Urban Artificial Intelligence and works with the ITU United for Smart Sustainable Cities initiative, under the Digital Transformation for People Centered Cities thematic. Since 2024, she has been working with the UNESCO Ethics of AI Unit, facilitating the design and implementation of ethical artificial intelligence in public administrations.

Leandry Junior Jieutsa

Is an urban innovation consultant and researcher on AI governance in cities. His career has been a journey through diverse roles in urban planning and development, underscored by his passion for integrating technologies into urban planning. He works as an independent consultant on urban innovation and digital governance. Leandry is also the founder of the Africa Innovation Network, a think tank dedicated to urban issues. He is currently enriching his knowledge as a PhD student at the UNESCO Chair in Urban Landscape of the University of Montreal, focusing on artificial intelligence governance in cities. His research investigates the impact of AI governance on people's well-being in cities. Through his research, he aims to support cities in building more people-centred smart cities through responsible AI governance.

Maria Pérez-Ortiz

Is Associate Professor at the AI Centre and Department of Computer Science at University College London (UK). She also acts as Senior Research Fellow at the Foreign, Commonwealth and Development Office of the UK government, providing advice on AI investments for

international development. In 2022, she co-founded the first MSc programme on AI for Sustainable Development, at the intersection of emerging AI technologies, sustainability and ethics. Pérez-Ortiz is also deputy of the UNESCO Chair in AI. Her latest policy report "Challenging systematic prejudices: an investigation into bias against women and girls in large language models" shows the extent to which language models show gender biases. Her current line of work is responsible AI and how these novel technologies could support policymakers in complex scenarios such as climate change.

Alexandra Vidal D'oleo

Is a researcher and project manager for CIDOB's Global Cities Programme at CIDOB (Barcelona Centre for International Affairs). Her research revolves around the digitisation of cities and the democratic and ethical processes in the development of urban artificial intelligence. Currently, she is part of the research team of the Global Observatory of Urban Artificial Intelligence (GOUAI) within the framework of the Cities Coalition for Digital Rights. She has a postgraduate degree in Territorial Planning with a focus on sustainable urbanism from the Universitat Politècnica de Catalunya and a degree in International Relations from Blanquerna - Universitat Ramón Llull, having completed part of her studies at L'École de Gouvernance et d'Économie in Rabat. Before joining CIDOB, she was a researcher and project manager at the Tanja Foundation, developing international cooperation projects between Morocco and Spain.

Accountability and transparency in urban AI

Shazade Jameson

Many urban public administrations see the potential of implementing artificial intelligence (AI) but feel unprepared for how to do so responsibly. While responsible AI approaches and frameworks are increasingly popular, much of these focus on private or industrial actors. Though there is increasing attention to the role of responsible AI in the public sector, there is much less guidance for local governments specifically. This chapter aims to narrow the gap by presenting definitions of accountability and transparency that include both narrow technical and broader sociopolitical perspectives. To facilitate accountability and transparency in the context of implementing AI by urban public administrations, there are two deceptively simple yet fundamental guiding questions for the design phase: "Should AI be used?" and "How is AI to be used?" After reflecting on what these questions mean for urban practitioners, the chapter presents a summary of existing policy mechanisms which can be adapted to work towards these aims and offers some lessons learned from previous research.

Privacy and data governance in urban AI

Josuan Eguiluz Castañeira and Carlos Fernández Hernández

The development of ethical artificial intelligence (AI) in Europe, as envisioned by the European AI Act, must include robust mechanisms for privacy and data management. In the context of public-urban environments, the processing of personal data through AI systems presents specific challenges that public authorities will need to address carefully. In light of this new legislative framework, the purpose of this article is to: (i) present the legal and ethical framework regulating the processing of personal data via AI systems in urban settings; (ii) outline key mechanisms to implement the principle of privacy; and (iii) examine the challenges associated with such data processing practices, providing a set of recommendations and best practices.

Fairness and non-discrimination in urban AI

Leandry Junior Jieutsa

Artificial intelligence (AI) is an emerging, disruptive and ambivalent technology. As part of its deployment, cities need to put various mechanisms in place to ensure that this technology has the least possible negative impact on people and communities. The aim is to ensure

that cities remain fair spaces that leave no one behind. This chapter, subdivided into three sections, formulates policy recommendations for integrating aspects of fairness and non-discrimination into the deployment of AI by cities. The first section discusses the notions of fairness and non-discrimination in urban settings and introduces the factors determining fair and non-discriminatory AI. The second section explores the opportunities and impacts of AI in cities. Finally, the third section proposes policy recommendations for fairer AI in cities. These recommendations take into account the different roles that cities can play in the deployment of AI as in-house solution developers, deployers and regulators. Cities need to be agile, relying on participation, local approaches, sociotechnical innovation, collaboration and so on.

Sustainability in urban AI

María Pérez-Ortiz

The chapter explores the potential of artificial intelligence (AI) to support the development of sustainable cities, addressing the social, environmental and economic dimensions of sustainability. As urbanisation accelerates globally, cities face increasing challenges in areas such as mobility, housing, pollution and resource management. AI holds promise for optimising urban infrastructure, reducing emissions and improving resource efficiency; however, its deployment also raises concerns about social equity, environmental impact and economic disruption. Sustainable AI is proposed as a framework for aligning AI's development and application with sustainability goals, ensuring it operates within ecological limits, promotes inclusivity, and supports equitable and circular economic growth. Key areas of focus include minimising AI's carbon footprint through energy-efficient practices, embedding fairness in AI-driven urban systems and ensuring transparent governance. The paper provides policy recommendations to guide AI deployment in urban settings, emphasising international collaboration, ethical governance and economic policies to foster resilience and inclusivity in the cities of tomorrow.

INTRODUCTION. ETHICAL URBAN AI IN PRACTICE

Tanya Álvarez
Researcher, Mobile World Capital Foundation Observatory

Marta Galceran-Vercher
Senior Research Fellow, Global Cities Programme, CIDOB

Alexandra Vidal D'oleo
Research Fellow and Project Manager, Global Cities Programme, CIDOB

Artificial intelligence (AI) is one of the most revolutionary technologies of our time and promises to completely transform society. This transformation is multilayered and ongoing in many spheres, the urban space being no exception. Furthermore, since the conception of the "smart city" paradigm, urban planners, tech companies and municipal policymakers are increasingly looking to technological advancements to solve the most pressing urban challenges our societies face. In this process, the deployment of algorithmic systems by local governments is widespread and shapes the process of city-making as we understand it.

City dwellers are only starting to see how AI as an integrated element of urban environments has a profound effect on the lived experience of cities and city-making itself.

"Urban AI" can be understood as the relationship between AI systems and the urban landscape. These systems, coupled with other technologies, are being embedded into all types of urban contexts: households, workplaces, public spaces and infrastructures. Moreover, the digitalisation of these urban experiences creates a hybrid environment where digital technologies play a role in mediating and augmenting the urban experience (Aurigi and De Cindio, 2008). City dwellers are only starting to see how AI as an integrated element of urban environments has a profound effect on the lived experience of cities and city-making itself.

When it comes to AI systems and automation, cities are an ideal testing ground for the deployment of these technologies. AI development and implementation requires a variety of resources which can be easily found in urban settings: a physical environment to act upon; access to a diversity of activities; copious amounts of high-quality data; and infrastructure and facilities (Cugurullo et al., 2023). In the past decade, there has been a surge of data-driven technologies that address urban challenges including infrastructure maintenance, personalised public services, health, transportation improvement, urban planning and efficient resource usage (Galceran-Vercher and Vidal, 2024).

As algorithmic technologies become increasingly commonplace, there is an urgent need for local administrations to be mindful of the responsible and ethical use of these systems.

An increasing number of municipal governments are aware of the benefits that AI brings to administration and delivery. AI systems are adopted in the hope that they alleviate the burden of routine, by automating bureaucratic tasks and thus allowing local governments to run more efficiently. They also seek to be more cost-effective by making smarter, data-driven decisions and freeing up local governments to better respond to the needs of their residents.

However, as algorithmic technologies become increasingly commonplace, there is an urgent need for local administrations to be mindful of the responsible and ethical use of these systems. Importantly, the growing weight that local governments carry in the global political arena, along with their potential impact on millions of lives, requires AI governance to consider the impact on individuals, communities and the environment.

With the advent of AI regulation, AI governance in cities has become of special concern for rights defenders, civil society organisations and minority urban populations as they have witnessed the potential pitfalls of the deployment of AI systems. For example, while AI enabled surveillance offers cities solutions regarding safety and security, traffic management or monitoring environmental factors, it has proven to be invasive and discriminatory towards certain sections of the population. This particular example, and others, raise the alarm on how efficiency gains from automation can come at great cost. Municipal administrations must therefore be aware of the ethical implications of the AI systems they seek to implement.

The challenge of operationalising ethical principles in urban AI

While cities may be concerned about the operational and technical benefits that AI promises, experts have argued that as sociotechnical systems, the impact of AI transcends the technical accuracy of the system itself. Consequently, local policymakers and administrators who only focus on the technical accuracy or fairness of a system fail to fully address the wider implications these systems may have when ensuring they are deployed responsibly and ethically. Implementing responsible AI goes beyond developing systems whose results are correct or reliable. Responsible and ethical AI emphasises the importance of ethics throughout the life cycle of a system, ensuring that algorithmic tools are aligned with democratic values and safeguard people's digital rights.

It is laudable that many cities have already started to implement and develop responsible AI policy mechanisms. New York City, for example, has introduced mandatory audits for hiring tools. In Finland, three cities have come together to promote AI transparency through algorithmic registers, making information accessible to residents. In another example, Toronto's police department has established a procurement policy for AI technologies. Furthermore, recent research by the Global Observatory on Urban AI (GOUAI) shows that cities worldwide routinely promote other policy mechanisms to incentivise responsible AI systems, such as the development of specific principles and guidelines; bans or

moratoria on specific high-risk algorithmic systems (e.g. real-time facial recognition systems); public algorithmic registries; impact assessments and audits; the establishment of external independent oversight bodies; or public procurement clauses that ensure compliance with human rights. These experiences can serve as a roadmap for other public sector players to understand what type of policy mechanisms work to develop responsible AI systems.

Still, while there are cities taking first steps to develop responsible AI practices, there is a growing need for city administrators and municipal policymakers to understand how urban AI is being developed and what best practices to put in place. The GOUAI, mentioned above, addresses this need. This is a joint project led by CIDOB, with the support of the cities of Barcelona, Amsterdam and London, the Cities Coalition for Digital Rights and UN-Habitat. To this end, the GOUAI's Atlas of Urban AI gathers cases of urban AI globally that adhere to six ethical principles: transparency and openness; privacy protection; fairness and non-discrimination; safety and cybersecurity; accountability; and sustainability. A recent publication based on the Atlas analysis (Galceran-Vercher and Vidal, 2024) revealed that with the growing trend of urban AI, there is a mismatch between cities that have adopted AI tools and those that have established policies or strategies to ensure that AI aligns with ethical principles.

Responsible and ethical AI emphasises the importance of ethics throughout the life cycle of a system, ensuring that algorithmic tools are aligned with democratic values and safeguard people's digital rights.

This CIDOB Monograph explores existing governance frameworks and specific policy mechanisms to operationalise concrete ethical principles and promote responsible urban AI on the ground. The aim is to create a useful document that inspires action and serves as a roadmap for other public sector players.

Structure of the publication

The first part of this publication comprises four chapters and examines how key ethical principles – accountability and transparency; privacy and data governance; fairness and non-discrimination; and sustainability – can be practically applied in urban settings through targeted policy mechanisms. **Shazade Jameson** argues that local governments lack clear guidance on advancing ethical urban AI in their jurisdictions and introduces two practical definitions of accountability and transparency that incorporate both technical and broader sociopolitical perspectives. Jameson argues that in order to foster accountability and transparency in urban AI implementation, two deceptively simple yet essential questions should guide the design phase: "Should AI be used?" and "How should AI be used?".

In the following chapter, **Leandry Junior Jieutsa** examines the factors that contribute to fair and non-discriminatory AI. He identifies two primary drivers of discrimination in AI systems: algorithmic biases and the use of AI technologies. Jieutsa offers policy recommendations aimed at creating fairer AI-powered cities, emphasising the need for local governments to adapt to their diverse roles as developers, deployers and regulators. He argues that cities must draw on participatory processes, localised approaches, sociotechnical innovation and cross-sector collaboration to ensure AI is deployed responsibly and equitably.

Recent research by the Global Observatory on Urban AI (GOUAI) shows that cities worldwide routinely promote other policy mechanisms to incentivise responsible AI systems.

Next, **Josuan Eguiluz Castañeira** and **Carlos Fernández Hernández** review mechanisms for a robust privacy and data management in AI deployment. They analyse the legal and ethical frameworks governing the processing of personal data by AI systems, with a particular focus on the European AI Act. They outline key mechanisms for implementing the principle of privacy in urban settings and explore the challenges associated with such data processing practices, offering a set of actionable recommendations. The authors emphasise that data governance must be central to urban AI strategies, prioritising the quality, relevance and protection of data sets used in AI systems. This includes conducting impact assessments to safeguard both personal data and fundamental rights, ensuring that citizens' privacy and security are not compromised.

Finally, **María Pérez-Ortiz's** chapter explores the potential of AI to contribute to the development of sustainable cities, addressing the social, environmental and economic dimensions of sustainability. The author argues that while AI offers significant promise, its deployment also raises concerns about social equity, environmental impact and economic disruption. In this regard, the sustainable AI framework provides a valuable tool for aligning AI's development and application with sustainability goals. It ensures that AI operates within ecological limits, fosters inclusivity and supports equitable and circular economic growth.

The second part of the publication features six case studies offering examples of local AI governance frameworks that cities worldwide have established, adopting concrete policy mechanisms to implement ethical urban AI in practice. Specifically, in this section **Alexandra Vidal D'oleo** explores the AI governance of the cities of Barcelona, Amsterdam, New York, San José, Singapore and Dubai.

The CIDOB Monograph wraps up with a concluding chapter in which **Marta Galceran-Vercher and Alexandra Vidal D'oleo** present a categorisation of policy mechanisms derived from the chapters, case studies and a literature review. This analysis examines the most widely used policy mechanisms and explores how they align with the different ethical principles. The authors also identify common trends and challenges cities encounter when trying to implement these ethical principles in practice, offering a set of general recommendations to address them.

References

Aurigi, A. and De Cindio, F. "Augmented Urban Spaces: Articulating the Physical and Electronic City". Farnham: Ashgate Publishing, 2008

Cugurullo F. *et al.* "Artificial Intelligence and the City. Urbanistic perspectives on AI". London: Routledge, 2023

Galceran-Vercher, M. and Vidal, A. "Mapping urban artificial intelligence: first report of GOUAI's Atlas of Urban AI". *CIDOB Briefings,* no. 56, 2024.

PART I. OPERATIONALISATION OF ETHICAL PRINCIPLES IN URBAN SETTINGS

Shazade Jameson
Senior Consultant, Digital Governance

1. Introduction

In this monograph, "ethical urban AI" means to implement responsible AI approaches within urban public administrations. Any discussion on responsible AI, therefore, must be attuned to the particular needs and situations of urban public administrations and their constituents.

Urban public administrations are a particular context; they are stewards of the public interest and operate very much at the local level. This makes for a particularly interesting and challenging environment, because urban public administrations are at once very close to local complexities and further away from national strategies. There is also incredible diversity in terms of size and capacities across administrations.

This means that while urban public administrations can draw on many insights from "ethical AI" and "responsible AI" approaches, repurposing these approaches can be limited because it requires a much broader perspective than many available resources suggest. Many "responsible AI" approaches fall under the umbrella of corporate governance, geared towards an industrial context: how can companies use AI for their products and services and do so responsibly? Urban public administrations have a different business model; presumably they focus first on the public interest.

There is an increasing attention to the role of responsible AI approaches in the public sector (see for example OECD, 2024). However, there is much less guidance for local governments specifically, particularly from a global perspective. This chapter aims to narrow this gap, by presenting definitions of accountability and transparency, situating these principles within the context of implementing AI by urban public administrations, and, finally, presenting a summary of existing policy mechanisms which can be adapted to work towards these goals.

2. Accountability and transparency principles

2.1. Accountability

There is an increasing attention to the role of responsible AI approaches in the public sector. However, there is much less guidance for local governments specifically, particularly from a global perspective.

Accountability is a concept with both broad and narrow definitions. Both of these types of definitions are important for local governments to consider, because of the organisation's position as a public body.

At its most basic, accountability is a form of relationship. The most widely accepted accountability theory in public administration (Bovens, 2007) states that accountability is a *relationship* between an actor and a forum, and the forum has the authority to say no. Accountability must specify *for what* and *to whom*. As a relationship, accountability is a social process that requires social engagement and a shared social understanding (Wieringa, 2020).

Accountability *for what* is often determined through procedural and substantive standards of public administration, and the ability to evaluate whether those standards have been met. Accountability *to whom* is extremely important for local governments, and it can be diverse sets of audiences. Who has what kind of accountability? The funder? The stakeholder? Impacted citizens? Because urban public administrations must consider the public interest, the pool of stakeholders and accountability bearers is much wide (Jameson *et al.*, 2021). Some use cases of AI may also touch on questions of political accountability, such as when the Dutch childcare benefits scandal led to the resignation of the government (Dachwitz, 2022; Amaro, 2021).

When local governments design public-facing use cases of AI, it is important for urban public administrations to engage with impacted communities from the design stage of the project (e.g. UN-Habitat & Mila Quebec AI Institute, 2022). Some responsible AI frameworks are narrow in scope and may be ill-equipped to meet the demands of a broader participatory process that is required in a public administration. In particular, the way that bias and inequalities become encoded in algorithms as a form of governance suggests that new forms of contestation and feedback need to be included in the organisational restructuring around AI governance (Taylor, 2021).

2.2. Transparency

Transparency with regard to AI is a layered principle. Like accountability, it has a long-established history as a mechanism in public administration, as well as in software engineering and computer science.

At a technical level, transparency is about disclosing information relative to an algorithmic system all along its life cycle. Transparency at these technical levels allows independent investigation and auditing of how models are used and their quality. This includes design purposes, data sources, hardware requirements, working conditions, expected system performance, and – importantly for algorithmic systems – the relationship between model variables and the architecture, as well as

characteristics of the data on which the model was trained. Transparency requires documenting the selection process for datasets, variables and the quality indicators for system development.

Data provenance (i.e. where the data comes from) and the quality of training data are very important to consider when implementing AI in public administrations. It is a significant limiting factor for the quality of algorithmic models and the primary source of bias in implementing AI in public administrations (UN-Habitat & Mila Quebec AI Institute, 2022; Longpre *et al*, 2023).

Transparency is an overarching principle for the field of explainable AI, which includes the ideas of explainability and interpretability. These concepts rapidly gained popularity as mechanisms for transparency and accountability at both a technical and socio-political level. The general purpose of the field is to open up the "black box" of closed algorithms which do not disclose the essence of their internal workings (Adadia and Berrada, 2018).

There are different approaches to the technical level of explainability; broadly, they fall into four categories (Wierenga, 2020). The first is explaining the model, such as providing clear instructions on what procedures algorithmic models follow and to what extent an algorithmic model can be explained in simple language to a non-expert human. The second is explaining the outcome, which means elaborating on the specific decisions made by algorithms and whether the mechanisms for making those decisions can be understood and evaluated or not. The third is inspecting the black box, which may take in a variety of techniques, such as visualising the inner workings of the algorithm. Finally, creating a transparent box is a design principle using explicit and visible predictors. Overall, the challenge for transparency at a technical level is that there is often a trade-off between interpretability and accuracy.

There is also an important socio-political layer to transparency beyond the technical level. This provides visibility on how algorithmic systems are used, which design choices are made by whom, and makes governance assumptions explicit. In these ways, transparency becomes an enabling condition for developing algorithmic accountability by providing ways forward for contestation.

There is also an important socio-political layer to transparency beyond the technical level. This provides visibility on how algorithmic systems are used, which design choices are made by whom, and makes governance assumptions explicit.

The two principles of transparency and accountability work together. Solutions for accountability often work on a principle of transparency, which must then be embedded within an institutional context that allows for accountability relationships to develop.

2.3. Working together

The two principles of transparency and accountability work together. Solutions for accountability often work on a principle of transparency, which must then be embedded within an institutional context that allows for accountability relationships to develop.

For example, algorithmic registers are tools for accountability. In practice, the way in which they work is to make information about algorithms and their use transparent, in a freely accessible register. (Jameson and Leal, 2022; Cath and Jansen, 2021). In this way, transparency is a vehicle which allows the evaluation of accountability in algorithmic system design.

Transparency, however, may be a necessary condition for accountability, but it is insufficient. For example, just because an algorithmic system is well-documented and transparent, it does not tell you why it was decided that this was evaluated as "good enough" for the purpose at hand, who decided this, and who was involved in the process. While transparency can function passively, accountability is more active: it includes not only how a system works, but why (Wierenga, 2020).

Given the amount of excitement and attention around the application of AI, there is a significant risk of techno-solutionism. Sometimes, a behavioural or social approach may be more suited to solve the problem at hand.

3. Implementing responsible AI for urban public administrations

When considering a responsible use of AI, there are two fundamental questions urban public administrations should ask: "Should AI be used?" and "How should AI be used?" Providing clear answers to these deceptively simple questions can create one of the most effective pathways towards transparency and accountability, because they make fundamental assumptions visible. This process also requires allocating time and resources.

3.1. Should AI be used?

AI is not neutral. Rather, AI embeds and reinforces the assumptions in its data and design. Without consciously designing AI towards a set of values that support the public interest, the structures of AI and its governance will embed values unconsciously, causing significant risks (e.g. UN-Habitat and Mila, 2022). The question of whether AI should be used is therefore not to be taken lightly.

For genuine accountability, the option to stop using AI must be on the table. "No" must remain a possibility. Otherwise, accountability becomes narrowed as a principle, alluded to as a virtue rather than as a functional relationship (Wierenga, 2020).

Second, the question of "should" is not only normative but also an operational question. Public administrations are seeking to achieve a particular purpose, and AI might be the best way to do that. Or it might not. Other data-driven or technological solutions may be more suited. In particular, AI and machine learning applications require a large amount of high-quality data, so when those conditions are not met, perhaps simpler data analytics may suffice.

Data-driven projects in municipalities often must deal with legacy infrastructure, old sensors, disconnected databases. This means that successful machine-learning applications in urban contexts require extended project discovery phases, sometimes up to 30-40% of project timelines. This time includes an investigation into the problem at hand, the current state of infrastructure and datasets, and which type of solution may be best suited. Budgets and stakeholder expectations need to provide space to accommodate this extended exploratory phase.

Given the amount of excitement and attention around the application of AI, there is a significant risk of techno-solutionism: the age-old challenge of a hammer looking for a nail. Sometimes, a behavioural or social

approach may be more suited to solve the problem at hand. Often, different types of solutions respond to different framings, which means the way we frame the problem sets the boundaries for the solution space. In other words, the way we think about the problem already defines the types of solutions we can create. This is not limited to AI but human-technology interactions more generally. A simple example is if the problem is that an elevator is too slow, rather than trying to optimise the speed of the elevator through mechanical engineering innovations, installing a mirror would mean people don't notice the boredom so much during the ride. An extended exploratory phase also allows stakeholders to ask the fundamental question: what is the problem we are trying to solve?

The extended exploratory phase includes significant local stakeholder collaboration, too. Successfully developing AI is almost always a collaborative affair and involves working with local universities, think tanks and businesses, especially considering the capacity gap that municipalities face. In Barcelona, for example, the machine learning algorithm developed for algorithmic-assisted decision-making in the intake procedure of the social services welcome centre was the result of significant collaboration between entities in order to make a locally relevant, bilingual algorithm (Jameson and Leal, 2022).

3.2. How is AI to be used?

While there are many different applications of AI in cities, within public administrations the tendency for using AI falls into two broad categories: automating existing processes, and data-driven predictions.

Automation means automating a part of existing bureaucratic processes or urban services. In this category, there is a logic or a process that already exists, and one part of that chain of events is going to be made faster or more efficient with the assistance of AI. When considering how to apply AI, the starting point is the current system.

Data-driven predictions are a different approach, because the starting point begins elsewhere: with a lot of data. Out of that data, data analysts will derive insights, and based on those insights, the administration designs new bureaucratic processes for urban services. Predictive modelling forms a new, data-driven logic in the administration (Kitchin, 2016).

While these two categories may use the same type of AI on a technical level (for instance, they may both use deep learning or image recognition techniques), the *way* that the AI is embedded within the processes of the city differs. The way AI is embedded within processes of the city changes the types of impacts that AI can have, and therefore changes how we think about accountability and transparency.

For example, when AI is being used to automate existing bureaucratic processes, existing review processes may be augmented with additional accountability mechanisms. For example, a quarterly review can be augmented with an additional impact assessment. Other process innovations may complement existing organisational habits in order to account for the lessons learned from embedding AI, such as feedback from the civil servants involved in the process, and citizen feedback.

When considering a responsible use of AI, there are two fundamental questions urban public administrations should ask: "Should AI be used?" and "How should AI be used?" Providing clear answers to these deceptively simple questions can create one of the most effective pathways towards transparency and accountability.

A socio-technical approach to AI recognises that what happens with an AI system is a result of the interaction between the technical and the social, between the system and how it is embedded within its context.

On the other hand, the use of data-driven predictions requires a slightly more complex approach to transparency and accountability because these are a new form of knowledge-making, which traditional public administrations are not equipped to process. In particular, predictive modelling changes the role of local expertise and where it is applied (Kitchin, 2016). Think of it like this: somebody with 20 years of experience walking on those corners may have a different perspective than what the data can read. Computational knowledge is different from experiential knowledge (van Ewijk and Baud, 2009), and algorithmic-assisted decision making may change the balance between the two.

Processes of accountability will require a dialogue between different ways of understanding, such as the difference between computational and experiential knowledge. How do we make sense of the current urban problem at hand? This "sense-making" or "meaning-making" is about deciding how we value different policy options and social results; and arguably it is something that AI is wholly dependent on humans to do (Tan, 2024). Thinking through and redesigning accountability processes and policy mechanisms presents an opportunity to evaluate the different types of meaning-making in play to ensure that the use of AI within public administrations is ethical.

4. Policy mechanisms

A socio-technical approach to AI recognises that what happens with an AI system is a result of the interaction between the technical and the social, between the system and how it is embedded within its context. That means in order to understand how an algorithmic system will function, it is important to understand how an algorithmic system interacts with its environment, and when which mechanisms can be most impactful.

An algorithmic system can be described by the "AI life cycle", which is a form of shorthand to describe the process of design, development and deployment. This is useful to understand because many of the risk management frameworks available are based on variations of this AI life cycle.

These are different options for policy mechanisms available at different stages of the AI life cycle. There are also overarching institutional governance mechanisms which occur throughout, and as a background to, the AI life cycle.

Framing and Design:

- **Impact assessments** usually take the form of a questionnaire to analyse potential social and ethical consequences before deployment. There are many variations of impact assessments, including Ethical IAs, Privacy IAs, Fairness IAs, etc. See for example UNESCO's Ethical Impact Assessment Tool.

- **Procurement clauses** are clauses in the contracts used by governments buying goods and services, in this case AI or AI-related services. While seemingly a bureaucratic formality, these can become a strategic lever

for public interest goals, for example by defining standards of auditability. See for instance the GovAI Coalition, spearheaded by the city of San Jose, which has created policy templates to be re-used by public administrations, including an AI FactSheet and a Vendor Agreement which binds vendors to requirements concerning performance, algorithmic bias, human oversight, and others. Eurocities is also developing procurement clause templates in line with the EU AI Act.

Development:

- **External algorithmic audits** are independent evaluations of an algorithmic system's workings to ensure compliance with ethical and legal standards. See for example the European Data Protection Board AI Auditing Checklist.

Deployment:

- **Algorithm registers and transparency standards** are publicly accessible lists that keep track of how public administrations are using algorithms or AI, in order to make that information accessible to the public and stakeholders. These repositories are based on a common scheme of metadata and information about the algorithm. See the Algorithm Transparency Standard, including the code schema used by nine European cities. A similar initiative is the UK's Algorithmic Transparency Recording Standard.

Policy and governance context:

- **Interdisciplinary governance oversight committees** bring together experts from a variety of fields, including law, ethics and social sciences, and representatives of affected communities to present a diverse set of perspectives in the oversight process. To be effective, these oversight boards must be independent and maintain a genuine veto power.

- **Participatory processes**, especially with affected communities, actively and meaningfully involve people at all stages of the AI life cycle, beginning from the framing and design rather than only *post hoc*. Through a more equitable process, these can help co-design more equitable outcomes.

- **Human-in-the-loop design** means humans remain involved as the key decision-makers throughout the points of a system to reduce errors and enabling overrides. While algorithmic systems are never fully removed from humans because all systems embed their design values (and many are corporately owned), the design approach remains useful to emphasise that humans should remain the final decision-makers.

5. Lessons learned

Previous CIDOB research (Jameson and Leal, 2022) explored case studies and experiences in municipal administrations applying accountability

and transparency mechanisms for urban AI. Specifically, the research explored the algorithm register in Amsterdam, the AI register in Helsinki, and a case of explainable machine learning developed for social services in Barcelona. This chapter highlights some of the recommendations and lessons learned for successful transparency and accountability initiatives.

Design:

- AI accountability and transparency initiatives worked well when these were framed as matters of the public interest, linking them to broader societal issues, and not just technical problems.

- Identifying priorities for the local municipalities leads to local definitions of success, which means that initiatives in one location can vary compared to another. In several cases, these variations were a response to events and news in the area.

- People will have different expectations of what an AI accountability initiative in the public administration can achieve. Successful projects required significant energy and had to have one designated "owner" of the project who was the primary reference person. That person spent a lot of time managing stakeholder expectations.

Process:

- Identify clear definitions that are understandable to all, non-expert civil servants. Key terms to ensure alignment are algorithm, transparency of what, when is it published, accountability to whom, and who is the product owner for what element of the project.

- Identify which organisational habits can be amplified with accountability processes. For example, existing quarterly financial report meetings were seen to be the moment that executives were already sitting around the table and could review additional technical innovations.

- Start small and iterate. Changes to how public administration works take time, and it works better when changes are made incrementally rather than in one fell swoop.

Capacity:

- All accountability initiatives required investments in capacity building to bring civil servants' education up to speed, as well as providing time to become familiar with new approaches

- Connect with knowledge-sharing networks, such as the Cities Coalition for Digital Rights, where experiences in adapting transparency and accountability mechanisms are shared and exchanged.

References

Adadi, A. and Berrada, M. "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)", *IEEE Access*, vol. 6, (2018), p. 52138–52160

Amaro, S. "Dutch government resigns after childcare benefits scandal" CNBC (January 2021) [Retrieved January 18, 2021]

Bovens, M. "Analysing and Assessing Accountability: A Conceptual Framework". *European Law Journal* vol. 13, no. 4 (2007), p. 447–68

Cath, C. and Jansen, F. "Dutch Comfort: The Limits of AI Governance through municipal registers". *arXiv*, (September 2021)

Dachwitz, I. "Childcare benefits scandal: Dutch government to pay million Euro fine over racist data discrimination". *Netzpolitik.org*, (January 2022)

Van Ewijk, E. and Baud, I. "Partnerships between Dutch municipalities and municipalities in countries of migration to the Netherlands; knowledge exchange and mutuality." *City-to-City Co-Operation*, vol. 33, no. 2 (2009), p. 218–26

Jameson, S. and Leal, A. "Transparency and accountability in urban artificial intelligence: Lessons from city initiatives". Global Observatory of Urban AI, CIDOB (2022)

Jameson, S., Taylor, L. and Noorman, M. "Data Governance Clinics: A New Approach to Public-Interest Technology in Cities". *SSRN* Scholarly Paper. Rochester, NY: Social Science Research Network (September 2021)

Rob, K. "The ethics of smart cities and urban science". *Phil. Trans. R. Soc. A,* vol. 374, no. 2083 (December 2016): 20160115

Longpre, S. *et al.* "The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing and Attribution in AI". *arXiv*, (November 2023).

OECD, "G7 Toolkit for Artificial Intelligence in the Public Sector" (2024)

Tan, V. "AI's meaning-making problem". Substack newsletter. The Uncertainty Mindset (Soon to Become Tbd) (blog), (May 2024)

Taylor, L. "Fairness and AI governance – responsibility and reality". *Global Data Justice* (blog), (April 2021).

UN-Habitat and Mila Quebec AI Institute. "AI & Cities: Risks, Applications and Governance", (2022)

Wieringa, M. "What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability", Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency. FAT* '20. New York (NY, USA): Association for Computing Machinery, (2020), p. 1–18

Josuan Eguiluz Castañeira
Legal Counsel, Adevinta

Carlos Fernández Hernández
Advisory Board Member, Global LegalTech Hub

1. Introduction

When the European Union (EU) began to devise its regulatory framework for artificial intelligence (AI) in 2018, from the outset it placed particular emphasis on this technology being "trustworthy". An AI system is deemed trustworthy if it complies with all applicable legislation and it is ethical and robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm (High-Level Expert Group on AI, 2019).

Consequently, the European approach on this matter incentivises the development and uptake of ethical and trustworthy AI across the EU economy, based on the principle that the technology should work for people and be a force for good in society (White Paper on Artificial Intelligence, 2020, § 6).

Given that the availability of data is essential to train algorithmic systems and that much of that data is personal, a component of ethical AI is that it must include privacy and data governance mechanisms (European Commission, Coordinated Plan on Artificial Intelligence 2021 Review). This requirement is fully incorporated into the European regulation on artificial intelligence (AI Act) of June 2024. It states that one of its purposes is to promote the uptake of human-centric and trustworthy AI (Article 1), while complying with the existing legal framework on data protection, which comprises – principally though not exclusively – the General Data Protection Regulation of 2016 (GDPR).

As several authors have noted (Almonacid Lamelas, 2024), the AI Act presents no small challenge to local governments, as they must adapt their processes, policies and strategies to meet the new requirements. But it is also an opportunity to improve their functioning, as well as the quality and trustworthiness of the AI-based services offered to citizens (ibid.). This explains the proliferation of "urban AI" systems, a concept that denotes "the collection, interpretation and analysis of urban data in order to support policy related decision-making and the development of solutions that are used, or could be used, in an urban context" (Galceran-Vercher, 2023).

Still, processing personal data in the public-urban sphere can raise specific problems, from the legitimacy of processing the data for a purpose that was not originally agreed to the need to carry out assessments of the impact on people's fundamental rights. These must clearly be taken into account by public bodies.

AI systems must also guarantee data protection throughout a system's entire life cycle. This includes the information initially provided by the user, as well the information generated about them over the course of their interaction with the system.

In light of the new legislative framework, the aim of this article is to (i) set out the legal and ethical framework that regulates personal data processing by AI systems in the urban sphere, particularly at the European level (AI Act); (ii) identify the main mechanisms for implementing the principle of privacy; and (iii) analyse the challenges that this type of data processing presents and offer a series of recommendations and good practices to minimise or rise to them.

2. AI ethics and privacy

Trustworthy AI must be ethical, and to do so it must, among other requirements, respect people's privacy. The AI Act sets the specific goal to "promote the uptake of human-centric and trustworthy AI". With that in mind, the common rules it lays down for high-risk AI systems must be consistent with the Charter of Fundamental Rights of the European Union (2000) and take into account both the European Declaration on Digital Rights and Principles for the Digital Decade (2022) and the ethics guidelines of the independent High-Level Expert Group on Artificial Intelligence (2019). According to these guidelines, in a context of rapid technological change,

"Trustworthiness is a prerequisite for people and societies to develop, deploy and use AI systems. Without AI systems – and the human beings behind them – being demonstrably worthy of trust, unwanted consequences may ensue and their uptake may be hindered, preventing the realisation of the potentially vast social and economic benefits that they can bring." (Introduction)

The trustworthiness of AI rests on three components, which must be present throughout the entire life cycle of the AI system:

1. It should be lawful, complying with all applicable laws and regulations;

2. It should be ethical, ensuring adherence to ethical principles and values; and

3. It should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm.

Ethics should therefore be a core pillar to ensure and scale trustworthy AI. This means that it is necessary to ensure alignment with some basic ethical norms, as well as with the measures laid down in the AI Act for the protection of fundamental rights.

Data protection is a fundamental right that is particularly affected by AI systems, and which is closely related to the principle of prevention of harm. That principle of prevention begins with adequate data

governance that covers the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy.

Those measures include AI systems having a privacy and data governance mechanism that takes in respect for privacy, quality and integrity of data, and access to data.

The AI Act gathers the general principle that the act fully complies with the EU's regulatory framework on data protection laid down in the GDPR.

AI systems must also guarantee data protection throughout a system's entire life cycle. This includes the information initially provided by the user, as well the information generated about them over the course of their interaction with the system (for example, the outputs the AI system generates for specific users or how they respond to particular recommendations). Digital records of human behaviour may allow AI systems to infer not only individuals' preferences, but also their sexual orientation, age, gender or religious and political views. To allow individuals to trust the data gathering process, it must be ensured that data gathered about them will not be used to discriminate against them unlawfully or unfairly.

Compliance with these requirements falls to the operators, particularly AI systems developers and those responsible for deploying the systems (who should ensure that the systems they use and the products and services they offer meet the requirements). Meanwhile, the people affected by the operation of an AI system shall have the right to be informed of that impact and, when applicable, lodge a complaint for breach of the AI Act (Articles 85 and 86).

2.1. Privacy and the European AI Act

Article 2(7) of the AI Act gathers the general principle that the act fully complies with the EU's regulatory framework on data protection laid down in the GDPR.

First, the harmonised rules laid down in the AI Act should apply across all sectors and should be without prejudice to existing EU law. It is important to point out, then, that the AI Act does not seek to affect the application of EU law governing the processing of personal data, including the tasks and powers of the independent oversight authorities that monitor compliance with those instruments. Similarly, nor does it affect the prior obligations of providers and deployers of AI systems in their role as data processors. In particular, the AI Act should not affect practices currently prohibited by EU law, including data protection law.

At the same time, the fact that an AI system is classified as high-risk should not be interpreted as indicating that its use is lawful under other acts of EU law or national law, for example on the protection of personal data. Any such use should continue to take place solely in accordance with the applicable requirements resulting from the Charter of Fundamental Rights, from EU secondary law and from national law.

Moreover, the AI Act does not provide for the legal ground for processing of personal data, including special categories of such data,

unless it is specifically otherwise provided for. Therefore, after the AI Act's entry into force, data subjects continue to enjoy all the rights and guarantees awarded to them by EU law, including those related to solely automated individual decision-making, such as profiling. The harmonised rules established under the AI Act should enable the exercise of the data subjects' rights and other remedies guaranteed under EU law on the protection of personal data and of other fundamental rights.

Finally, in order to facilitate compliance with EU data protection law, in specified conditions the AI Act provides the legal basis for the providers (and prospective providers) in the regulatory sandbox to use personal data collected for other purposes to develop certain AI systems in the public interest.

For cities, ensuring their AI systems comply with regulations such as the GDPR or the AI Act throughout the system's entire life cycle is crucial to safeguard citizens' rights and maintain public trust.

3. Policy mechanisms for implementing the principle of privacy in the urban environment

Privacy and data protection in the implementation of urban AI requires the adoption of specific policy mechanisms. These mechanisms allow cities to comply with existing regulations and ensure that AI is deployed ethically and responsibly, respecting citizens' rights. Below, we spotlight and explain the main policy mechanisms for implementing this ethical principle.

a) Ensuring legal compliance

Compliance with regulation is an essential ethical requirement of privacy and data protection in public authority implementation of AI systems in urban environments. For cities, ensuring their AI systems comply with regulations such as the GDPR or the AI Act throughout the system's entire life cycle is crucial to safeguard citizens' rights and maintain public trust. This includes adherence to key requirements such as the quality and integrity of the data used, its relevance in light of the domain in which the AI systems will be deployed, its access protocols and the capability to process data in a manner that protects privacy (High-Level Group of Experts on AI, 2018).

Indeed, these requirements are present as specific obligations in the AI Act itself, purposely designed for high-risk cases such as AI systems for remote biometric identification – e.g. the ABIS program (Pascual, 2024) – or those used to assess a natural person's eligibility for essential public assistance services and benefits – e.g. the Syri case (Digital Future Society, 2022).

b) Risk management and data governance systems

The AI Act includes specific obligations (Articles 9 and 10) closely linked to the principle of privacy and data protection. Article 9 focuses on the creation of a risk management system capable of identifying, documenting and mitigating the risks associated with the use of AI in cities. These risk management systems should establish continuous

iterative processes planned and run throughout the entire life cycle of AI technologies which, of course, will require regular systemic review and updating. In fact, not only does it mean assessing possible risks before the introduction into the market or the entry into service of these AI systems, but also setting up and/or supervising the functioning of a post-market monitoring system to manage emerging risks (Articles 17(1) h, 26(5) and 72 of the AI Act).

Data governance regulated in Article 10, meanwhile, requires the training, validation and testing data sets used in high-risk AI systems to be subject to data governance and management practices appropriate for its intended purpose. The practices to be implemented by cities to ensure effective and lawful data governance should focus on matters such as data collection processes and the origin of data; the purpose of the data processing; an assessment of the availability, quantity and suitability of the data sets needed; examination of possible biases that might affect the health, safety or fundamental rights of persons, and so on.

c) Impact assessments

Article 35 of the GDPR requires controllers (e.g. local authorities) to carry out a data protection impact assessment (DPIA). This assessment shall be carried out when a type of processing, given its nature, scope, context or purposes (in particular using new technologies), is likely to result in a high risk to the rights and freedoms of natural persons (AEPD, 2018; Article 29 Working Party, 2017; Friedwald et al., 2022). This preventive approach is vital in urban environments to anticipate possible data protection vulnerabilities and take the necessary steps to remedy them in a timely manner.

Likewise, for high-risk AI systems, Article 27 of the AI Act introduces the obligation to carry out a fundamental rights impact assessment (FRIA) (Government of the Netherlands, 2022; Danish Institute for Human Rights, 2020) to complement the DPIA. This assessment aims to identify the specific risks to the rights of individuals likely to be affected and establish measures to be taken in the event of a materialisation of those risks (Recital 96 AI Act). It is worth noting that the impact assessments (Manzoni et al., 2022) should focus not only on return on investment, but also on the sustainability and ethical impact of technology, addressing financial, human and environmental aspects (OECD, 2024).

d) Auditing

Having said this, it will be necessary to demonstrate to authorities, stakeholders and citizens that there is compliance with the law and all its specific implementation requirements. Accordingly, internal and external audits shall be carried out and certification shall be obtained to verify that systems operate within the established legal frameworks. To that end, European cities, for example, should carry out conformity assessments (Article 43 AI Act) in order to ensure and demonstrate compliance with the requirements associated with high-risk systems,

in line with the harmonised standards published in the *Official Journal of the European Union* (Article 41 AI Act). They should also follow the common specifications established by the European Commission, thus ensuring standardised and safe implementation of AI systems (e.g. ISO certifications).

AI audits are considered a fundamental governance mechanism to ensure that the deployment and operation of AI systems comply with established legal regulations and ethical and technical standards (Fernández and Eguiluz, 2024). Generally speaking, these audits should be carried out by independent and competent bodies. The auditing process includes methodologies that incorporate ethical impact assessments (UNESCO, 2024; CEN-CENELEC, 2017), ensuring that AI systems behave responsibly and that their impacts on society and on individuals are properly monitored and mitigated. It is, however, recommendable to consider AI audits from a multidisciplinary (legal, technical and ethical) perspective (Mökander, 2023). Thus, proposals such as "algo-scores" have arisen to classify and assess in an accessible manner an algorithmic system's level of conformity on matters such as ethical compliance, AI governance, the equity of the model and its subsequent monitoring, taking a similar approach to energy efficiency labelling (Galdon Clavell, 2024).

e) Algorithm repositories and AI systems registers

Lastly, it is important to remember the importance of public algorithm repositories and AI systems registries (Article 49 AI Act) that promote transparency in automated decision-making in the public sector and play a crucial role in protecting privacy and personal data. By making details of how these systems are designed, deployed and operated accessible, repositories and registries enable citizens and organisations to understand how and for what purposes their personal data is used in these processes. These repositories also include information about the data sources used and oversight mechanisms, which is essential to assess the impact on individual privacy and ensure that data protection measures are effective (Gutiérrez and Muñoz-Cadena, 2024).

4. Challenges and recommendations

Inadequate data management is one of the chief limitations when it comes to deploying AI in the public sector. As is lack of access to sufficient volumes of high-quality data. This problem is exacerbated by unsatisfactory sharing of data across organisations owing to the absence of unified standards and underdeveloped data governance. In addition, distrust of AI systems compounds these challenges. Scattered laws and insufficient knowledge of the impacts of AI also form significant barriers (Manzoni et al., 2023). Likewise, increasing cyberattacks have led to the NIS 2 Directive (2022) boosting the level of security and legal responsibility for administrators. In 2023, the public administration was one of the most affected sectors, registering 19% of reported incidents, with a marked rise in ransomware and DDoS attacks (ENISA, 2023).

The complex regulatory landscape also presents a significant challenge. Interaction between urban regulations at European, national and local level creates a web of rules that hampers effective AI deployment in cities. Urban legislation and specific regulations in each municipality should align with European laws such as the Interoperable Europe Act (2022), which seeks to improve the interoperability of digital public services (Tangi et al., 2023).

Another major limitation is the lack of experience and technical knowledge in local administrations, which hinders proper implementation of AI. The general shortage of professionals in the field, coupled with growing competition for talent, present a significant barrier for cities that are trying to develop and deploy these systems effectively (OECD, 2024).

Additionally, the mass collection of personal data, which is required to train these systems, may infringe a citizen's right to control their data as it may be sensitive or managed inappropriately. Meanwhile, AI applications like those used in policing can intensify mass surveillance and compromise individual privacy still further (Véliz, 2020; Agarwal, 2018; Dwivedi et al., 2019).

In order to overcome these barriers, it is essential to promote innovation mechanisms such as regulatory sandboxes (Madiega, 2022) that allow cities to experiment with AI in a controlled environment while guaranteeing regulatory compliance (Tangi et al., 2023). Likewise, coordination between national authorities (in Spain's case, the Spanish Artificial Intelligence Oversight Agency, AESIA) and European bodies (the European AI Office) is crucial to ensure that AI systems comply with existing regulations and are deployed safely and responsibly.

It is essential to promote innovation mechanisms such as regulatory sandboxes that allow cities to experiment with AI in a controlled environment while guaranteeing regulatory compliance.

Interoperability and collaboration are equally crucial. Initiatives such as the SALER – a rapid alert system used in the autonomous community of Valencia to prevent corruption in the administration – show how AI can be used effectively to improve governance processes (Digital Future Society, 2023). Likewise, it is essential that public funding is conditional on the various administrations making specific outputs available (e.g. generating public data sets) (European Commission, 2022). To this end, the European Commission published in its Implementing Regulation (EU) 2023/138 a list of specific high-value data sets that should be available for free re-use, highlighting the potential of public data to benefit society, the environment and the economy (European Commission, 2022). In addition, access to multilingual data to train local AI models that reflect the specific characteristics of each region (OECD, 2024) and the collection of AI use cases in the public sector at European level (European Commission, 2021) will improve AI systems' effectiveness and equity while providing a valuable source of information on how these technologies are being implemented in different contexts.

5. Conclusions

The general data protection framework in the EU now rests on a set of principles that the EU's administrative and judicial bodies are aware of and solidly interpret. AI, however, poses specific problems

of a technological and legal nature that are at a nascent moment of knowledge and treatment.

Numerous studies, experiences and clarifications shall be still necessary, then, to provide them with a legal framework that ensures the proclaimed purpose that AI should be human centred, a tool for people and have the ultimate goal of improving their well-being.

The introduction of specific policy mechanisms is essential to ensure that cities use AI systems in a manner that is ethical and respects citizens' rights. Compliance with regulations such as the GDPR and the AI Act is crucial to safeguard privacy and personal data in urban environments. Similarly, it is paramount that cities establish risk management systems that iteratively address contingencies associated with the AI's entire life cycle, including regular reviews and external audits to ensure regulatory compliance.

Data governance, too, must be at the heart of urban AI strategies. Cities must implement sound data governance and management practices, focusing on the quality, relevance and protection of the data sets used in AI systems. This includes conducting impact assessments both for the protection of personal data and for fundamental rights, ensuring that the technology deployed does not breach citizens' privacy or security.

Ultimately, achieving human centred AI will require a joint effort among those responsible for developing public policies, academic institutions and the private sector, who must work together to ensure that AI systems implemented by cities align with fundamental values and ethical principles.

As stated, the future of smart cities will be marked by the synthesis of multiple technologies aimed at satisfying the intricate mosaic of human needs. This convergence will require precise optimisation of the technologies applied to ensure that the digitalisation of urban spaces conforms to sustainable and equitable practices, as well as attentiveness to the ethical dimensions involved in these innovations. It is therefore imperative that the integration of AI into the heart of smart cities abides by principles that protect privacy, security and inclusion. As Zhenjun et al. (2023) say: "Ensuring that the benefits of smart city developments are equitably shared will be essential in avoiding societal fractures and fostering an environment where technology serves as a bridge to a more enlightened, harmonious urban life."

References

Agencia Española de Protección de Datos (AEPD). "Gestión del riesgo y evaluación de impacto en tratamientos de datos personales". Madrid, 2018 [accessed: 2 September 2024]

Agencia Española de Protección de Datos. "Guía sobre protección de datos y administración local". Updated 2023 [accessed: 2 September 2024]

AI Ethics Impact Group. "From Principles to Practice: An interdisciplinary framework to operationalise AI ethics". 1 April 2020 [accessed: 2 September 2024]

Alan Turing Institute. "Urban analytics" [accessed: 2 September 2024]

Alan Turing Institute. "Why the public sector needs to know about AI ethics (and how we're helping)". 2 November 2023 [accessed: 2 September 2024]

Almonacid Lamelas, V. "Reglamento (europeo) de Inteligencia Artificial: impactos y obligaciones que genera en los Ayuntamientos". El Consultor de los Ayuntamientos, LA LEY, 15 July 2024 [accessed: 2 September 2024]

Article 29 Working Party. "Guidelines on Data Protection Impact Assessments (DPIA) and determining whether processing is "likely to result in a high risk" for the purposes of Regulation 2016/679", Brussels, 2017 [accessed: 2 September 2024]

Burbano, L. (1). "Privacy protection in smart cities: How are they taking care of citizens' most precious information?" *Tomorrow.city*, 23 January 2024 [accessed: 2 September 2024]

Burbano, L. (2). "AI urbanism: risks and benefits of a seemingly unstoppable movement". *Tomorrow.city*, 22 February 2024 [accessed: 2 September 2024]

Canda, J. "AI in Urban Planning and Smart City Development". *Medium*, 7 April 2024 [accessed: 2 September 2024]

CEN-CENELEC. "Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework". CWA 17145-2, June 2017 [accessed: 2 September 2024]

Centro Latinoamericano de Administración para el Desarrollo (CLAD). "Carta Iberoamericana de Inteligencia Artificial en la Administración Pública". 20 November 2023 [accessed: 2 September 2024]

Danish Institute for Human Rights. "Guidance on Human Rights Impact Assessment of Digital Activities". Copenhagen, 2020 [accessed: 2 September 2024]

Digital Future Society (1). "Algorithmic discrimination in Spain: limits and potential of the legal framework", August 2022 [accessed: 2 September 2024]

Digital Future Society (2). "El acceso digital en las ciudades, entendido como algo más que un derecho fundamental", June 2023 [accessed: 2 September 2024]

Digital Future Society (3). "El uso de algoritmos en el sector público en España: cuatro estudios de caso sobre ADMS", February 2023 [accessed: 2 September 2024]

European Commission (1). "Selected AI cases in the public sector (JRC129301)". Joint Research Centre (JRC), 2021 [accessed: 2 September 2024]

European Commission (2). "Revisión de 2021 del plan coordinado sobre la inteligencia artificial", Anexos de la Comunicación de la Comisión al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Fomentar un planteamiento europeo en materia de inteligencia artificial, COM(2021) 205 final, 2021 [accessed: 2 September 2024]

European Commission (3). "Opportunities and challenges of artificial intelligence technologies for the cultural and creative sectors". Publications Office of the EU, Luxembourg, 2022 [accessed: 2 September 2024]

European Commission (4). "Second Report on the application of the General Data Protection Regulation", COM(2024) 357 final, Brussels, 25 July 2024 [accessed: 2 September 2024]

European Data Protection Board (EDPB). "Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework", 16 July 2024 [accessed: 2 September 2024]

European Union Agency for Cybersecurity (ENISA). "Threat Landscape 2023", October 2023 [accessed: 2 September 2024]

Fernández, C. and Eguiluz, J. A. "Diez puntos críticos del Reglamento europeo de Inteligencia Artificial". *Diario LA LEY*, No. 85, Sección Ciberderecho, 28 June 2024 [accessed: 2 September 2024]

Friedewald, M *et al*. "Data Protection Impact Assessments in Practice: Experiences from Case Studies". *Computer Security, ESORICS 2021 International Workshops*, February 2022, pp. 424-443 [accessed: 2 September 2024]

Galceran-Vercher, M. "Trustworthy Cities: Ethical Urban Artificial Intelligence". The GovLab, Course "AI Ethics, Global perspectives", December 2023 [accessed: 2 September 2024]

Galceran-Vercher, M. and Vidal, A. "Mapping urban artificial intelligence: first report of GOUAI's Atlas of Urban AI". Global Observatory of Urban Artificial Intelligence (GOUAI), 2024 [accessed: 2 September 2024]

Galdon, G. "AI Auditing. Proposal for Algo-scores". EDPB, 27 June 2024 [accessed: 2 September 2024]

Ghisleni, C. "Artificial Intelligence and Urban Planning: Technology as a Tool for City Design". *ArchDaily*, 8 February 2024 [Last accessed: 2 September 2024]

Government of the Netherlands. "Impact Assessment Fundamental Rights and Algorithms", 31 March 2022 [accessed: 2 September 2024]

Gutiérrez, J. D. and Muñoz-Cadena, S. "Algorithmic transparency in the public sector: A state-of-the-art report of algorithmic transparency instruments". *Global Partnership on Artificial Intelligence*, May 2024 [accessed: 2 September 2024]

Hadec, J., Di Leo, M. and Kotsev, A. "AI generated synthetic data in policy applications". *Science for Policy Brief*, European Commission, Joint Research Center, 2024 [accessed: 2 September 2024]

High-Level Expert Group on AI. "Directrices éticas para una IA fiable", April 2019 [accessed: 2 September 2024]

Imdat As, P.; Basu, P. and Talwar, P. *Artificial Intelligence in Urban Planning and Design. Technologies, Implementation, and Impacts*. Amsterdam: Elsevier, 2022.

Madiega, T. and Van De Pol, A. L. "Artificial intelligence act and regulatory sandboxes". European Parliamentary Research Service (EPRS), PE 733.544, June 2022 [accessed: 2 September 2024]

Manzoni, M. *et al.* AI Watch. Road to the adoption of Artificial Intelligence by the public sector. Publications Office of the EU. Luxembourg, 2022. JRC129100 [accessed: 2 September 2024]

Mökander, J. "Auditing of AI: Legal, Ethical and Technical Approaches" *Digital Society*, vol. 2, article no. 49, 2023 [accessed: 2 September 2024]

Organisation for Economic Co-operation and Development (OECD) (1). "Advancing Accountability in AI: governing and managing risks throughout the lifecycle for trustworthy AI", *OECD Artificial Intelligence Papers,* no. 349, 23 February 2023 [accessed: 2 September 2024]

Organisation for Economic Co-operation and Development (OECD) (2). "Governing with artificial intelligence: Are governments ready?", *OECD Artificial Intelligence Papers*, no. 20, June 2024 [accessed: 2 September 2024]

Pascual, M. G. "La Policía española ya usa en sus investigaciones un sistema automático de reconocimiento facial". *El País*, 28 May 2024 [accessed: 2 September 2024]

Pellegrin, J., Colnot, L. and Delponte, L. "Artificial Intelligence and Urban Development". Research for REGI Committee, European Parliament, Policy Department for Structural and Cohesion Policies, Brussels, 2021 [accessed: 2 September 2024]

Tangi, L. *et al.* "Artificial Intelligence for Interoperability in the European Public Sector: an exploratory study", Publications Office of the EU, Luxembourg, 2023 [accessed: 2 September 2024]

Tangi, L. *et al*. "AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector", Publications Office of the EU, Luxembourg, 2022 [accessed: 2 September 2024]

Timan, T., Van Veenstra, A. F. and Bodea, G. "Artificial Intelligence and public services". Policy Department for Economic, Scientific and Quality of Life Policies, PE 662.936, July 2021 [accessed: 2 September 2024]

United Nations (UN). "Recommendation on the Ethics of Artificial Intelligence", 23 November 2021 [accessed: 2 September 2024]

Véliz, C. *Privacidad es poder: Datos, vigilancia y libertad en la era digital*. Madrid: Debate, 2021.

Verhulst, S. G. "Are we entering a Data Winter? On the urgent need to preserve data access for the public interest" [accessed: 2 September 2024]

Yan, Z. *et al*. "Intelligent urbanism with artificial intelligence in shaping tomorrow's smart cities: current developments, trends, and future directions". *Journal of Cloud Computing*, 18 December 2023 [accessed: 2 September 2024]

Leandry Junior Jieutsa
Researcher on AI Governance in cities, UNESCO Chair in Urban Landscape

1. Introduction

Fairness and non-discrimination are core values of urban AI in people-centred smart cities. Increasing discussion among researchers and policymakers testifies to the growing importance of addressing bias and discrimination in AI systems. Fairness derives from moral judgment, i.e. the process by which individuals determine what is morally right or wrong (Weinkauf, 2023). Although AI offers many advantages for cities, its deployment puts the quest for a fair city to the test by creating or reinforcing discrimination and inequalities. Thus, integrating fairness and non-discrimination principles into the urban AI life cycle is crucial to ensure the well-being of individuals and communities in smart cities. Nevertheless, operationalising this principle remains complex and ambiguous. To achieve this, cities need to articulate their various roles in AI governance, whether they are developers of internal solutions, responsible for the deployment of external systems or regulators . This requires the adoption of a variety of mechanisms, including socio-technical innovation, the establishment of local standards for fairness in AI and procurement standards. In addition, urban legislation must be introduced to protect the most vulnerable and guarantee citizens the exercise of their digital rights. However, these measures require resources, which cities can mobilise by promoting cooperation and networking.

2. A fair AI system is bias-free and used responsibly

Fairness and non-discrimination are complex and critical concepts in contemporary society (Barocas et al., 2023a). According to the Cambridge Dictionary, "fairness" refers to the quality of treating individuals equally in a manner that is just or reasonable. It respects people both as individuals and as members of society. Three primary elements, articulated in distributive and socio-relational dimensions, constitute this concept that pertains to individuals or groups (Barocas et al., 2023b): fair equality of opportunity, right to justification, and equality in relationships (Giovanola and Tiribelli, 2022). A fair society necessitates considering each individual or group of individuals according

to their specific characteristics and circumstances to ensure equitable treatment and outcomes (Giovanola and Tiribelli, 2022; Lyu et al., 2023). Thus, it incorporates the notion of non-discrimination, which implies that no one should be excluded. Vulnerable individuals or groups are most susceptible to discrimination.

Operationalising this principle remains complex and ambiguous. To achieve this, cities need to articulate their various roles in AI governance, whether they are developers of internal solutions, responsible for the deployment of external systems or regulators.

The emergence of disruptive technologies such as AI challenges the dimensions of fairness. Two main factors are involved in the context of discrimination in connection with AI, namely algorithmic biases and the utilisation of AI-based systems (Ferrara, 2023; O'Neil, 2016; Wachter et al., 2021).

The first factor, algorithmic bias, distorts the original training data or the AI algorithm, leading to skewed and potentially detrimental results (Holdsworth, 2023). These biases reduce the accuracy and potential of AI with varying degrees of impact depending on the application. There are two main categories of bias in AI: automation bias and bias by proxy (Barocas et al., 2023a; González-Sendino et al., 2023). Automation bias is the large-scale propagation through AI system processes of social and cultural biases deeply embedded in historical training data used to fuel the AI system. This category includes human bias, data bias, learning bias and deployment bias. Bias by proxy happens when unintentional proxies for protected variables (e.g. gender, race) allow biases to be inferred, despite efforts to exclude them from training data.

The second factor is the utilisation of AI-based systems. Indeed, when employed for profiling or social control, systems infringe upon digital rights (Calzada, 2021; Cugurullo et al., 2022). By collecting and utilising personal information, facial recognition technologies, for instance, violate the privacy and personal data of citizens (UN-Habitat, 2023). Digital rights are interpreted as existing human rights that must be protected in the context of digital technologies, as physical and digital spaces are increasingly intertwined (UN-Habitat, 2020).

Algorithmic fairness is predicated on interrelated variables (Weinkauf, 2023). An automated decision system is considered fair when it does not rely on sensitive data such as gender or religion, does not disadvantage minorities, and is utilised responsibly.

3. The AI dilemma: balancing between opportunities and impacts of AI systems in cities

Historically, urban planning has contributed to creating and reinforcing different forms of urban inequalities and discrimination (Fainstein, 2009; Hall, 2014). The most affected populations are notably minorities and the most vulnerable, which vary according to context. Consequently, numerous concepts have emerged, such as Henry Lefebvre's "right to the city" or the "just city" (Fainstein, 2009; Fincher and Iveson, 2012; Harvey and Potter, 2009; Lefebvre, 1968). These concepts aim to make cities more equitable, particularly through access to urban services and opportunities, for an improved quality of life.

The emergence of AI challenges the just city by bringing opportunities for more inclusive cities while also creating and reinforcing different

forms of inequalities and discrimination. Indeed, AI systems possess the capability to filter and process substantial volumes of data connected to extensive networks and the urban environment. Consequently, they can enable complex decisions to be made autonomously or semi-autonomously (Marvin et al., 2022; Sherman, 2023; Yigitcanlar et al., 2021). Explainable AI (XAI) methodologies can assist municipalities in comprehending the calculation of equity and its improvement (Lyu et al., 2023). The implementation of AI facilitates enhanced citizen-municipality engagement and optimises service delivery, particularly for the most vulnerable populations.

For instance, deep learning tools enhance spatial data management to optimise service delivery in disadvantaged neighbourhoods in Durban, South Africa. Generative AI facilitates participatory planning processes by generating urban scenarios in real time, thus enabling more inclusive urban planning that incorporates diverse perspectives. Furthermore, municipal chatbots, such as those implemented in Helsinki, Finland, or Saint-Lin-Laurentides, Canada, automate citizen interaction. This improves the management of public services, particularly for individuals unfamiliar with often complex administrative procedures, or those who face difficulties in accessing services in person.

Unlike traditional forms of discrimination, discrimination automated by algorithms is more abstract or opaque and unintuitive, subtle, intangible, difficult to detect and large-scale.

However, as previously stated, AI systems and the emphasis on the economic competitiveness of cities challenge the just city by producing unfair and discriminatory outcomes. Moreover, unlike traditional forms of discrimination, discrimination automated by algorithms is more abstract or opaque and unintuitive, subtle, intangible, difficult to detect and large-scale (Kleinberg et al., 2018; O'Neil, 2016; Sanchez et al., 2024; Wachter et al., 2021).

For example, tax algorithms targeting "foreign-sounding names" and "dual nationality" led to thousands of racialised families being falsely accused of fraud in the Netherlands. Globally, predictive policing systems, like Clearview AI, raise privacy concerns while reinforcing bias (Dauvergne, 2022; O'Neil, 2016). In 2021, *Forbes* reported algorithmic bias in mortgage applications, with 80% of Black applicants denied. Similarly, *The Markup* (2021) found applicants of colour were 40-80% more likely to face loan denials, underscoring the discriminatory impact of AI.

Furthermore, the concentration of wealth in large cities, due to urban AI, leads to urban gentrification (Sanchez et al., 2024). Access, particularly to housing, for low-income populations is becoming increasingly difficult, if not impossible. Urban AI deployment policies thus contribute to reinforcing asymmetries between territories and urban inequalities.

AI systems therefore have significant impacts on cities and societies. This ambivalence raises the need for effective governance. Additionally, due to its opacity and the scale of its impact, it becomes challenging for affected individuals to defend themselves or assert their rights. This calls into question the right to non-discrimination enjoyed by citizens because algorithmic decision-making systems disrupt traditional legal remedies and procedures for detecting, investigating, preventing and correcting discrimination (Wachter et al., 2021).

4. Policy recommendations

Due to its opacity and the scale of its impact, it becomes challenging for affected individuals to assert their rights […] because algorithmic decision-making systems disrupt traditional legal remedies and procedures for detecting, investigating, preventing and correcting discrimination.

According to UNESCO recommendations, AI actors must adopt an inclusive approach aimed at rendering the benefits of AI technologies available and accessible to all, taking into account the specific needs of different groups (UNESCO, 2023). At the city level, the implementation of fairness and non-discrimination in urban AI systems necessitates the articulation of the diverse roles that municipalities assume as developers of in-house solutions (albeit relatively infrequently due to financial and technical constraints), deployers and regulators. Enhancing the equity of AI systems additionally entails consideration of their entire life cycle, addressing various aspects throughout the design, development and implementation processes. Furthermore, the provision of effective solutions to disparities in AI system outcomes commences with the identification of their underlying causes.

4.1. General recommendations:

- **Define a strategy:** Cities must implement AI strategies that are structured around the principles of fairness and non-discrimination. These strategic documents enable cities to establish a robust foundation and conduct a precise assessment of their AI-related objectives. This approach is essential for planning the integration of AI to maximise its benefits while mitigating potential risks. These strategies should be developed through a participatory process and accompanied by action plans that delineate concrete measures to ensure the equitable integration of AI that leave no one behind.

- **Establish risk levels according to applications:** Cities must identify high-risk AI applications within their jurisdictions, taking into account existing disparities and inequalities in the territory. The identification of these high-risk applications should be followed by the implementation of protective mechanisms. Applications related to essential social services should be classified as high-risk and prohibited from operating with full autonomy. For instance, the City of San Jose has implemented an AI registry structured around a rigorous assessment of AI systems. This process involves a risk analysis, followed by a more comprehensive impact assessment, depending on the level of risk, all documented via an "Impact Sheet" and an "AI Fact Sheet".

4.2. Specific recommendations for cities as developers of in-house solutions

- **Emphasise inclusive socio-technical innovation.** Incorporate diverse non-technical stakeholders throughout the AI life cycle. According to UN-Habitat, this AI life cycle comprises five phases: framing, design, implementation, deployment and maintenance. If decisions in these various stages are predominantly made by technical actors or homogeneous groups, there is a significant risk that their biases will be integrated into the AI system. This risk is particularly pronounced if the tool is subsequently applied or generalised to broader population segments. Local governments must place greater emphasis on interdisciplinarity and multidisciplinarity that integrates social groups into the life cycle of urban AI.

- **Implement fairness techniques,** including: preprocessing data (which involves identifying and addressing biases in the data prior to model training); model selection (which focuses on utilising model selection methods that prioritise fairness); and postprocessing decisions (which involves adjusting the output of AI models to mitigate bias and ensure fairness) (Ferrara, 2023).

- **Enhance diversity in database construction across three dimensions: teams, data and models.** Establishing diverse, interdisciplinary teams and implementing ongoing training in fairness and ethics are crucial for minimising biases. Regarding data, enhancing the collection of sensitive attributes (e.g. sex, race, ethnicity) and documenting data-related decisions promotes transparency and facilitates addressing real-world inequalities. For models, providing open access to the community for testing, ensuring transparent documentation, and utilising explainable AI (XAI) can aid in identifying and mitigating biases, thereby ensuring equitable outcomes (González-Sendino et al., 2023).

- **Integrate compensatory correlation in AI systems.** As indicated by Giovanola and Tiribelli (2022), ensuring fair equality of opportunity in AI systems cannot be limited to eliminating discriminatory biases in the training data. Urban AI systems should be designed to consider existing inequalities in their context and incorporate mechanisms to compensate for them. For instance, in a city where disparities exist among communities or social groups, urban AIs must account for these disparities and implement compensatory measures. This may manifest in the form of personalised content, for example.

- **Integrate mitigation techniques in the AI life cycle.** Neutralise discriminatory effects in the data during the pre-training phase through methods such as resampling (altering the size of the data set that affects the distribution without transforming the data), fair representation (achieved by eliminating information that can associate an individual with a protected group), and re-weighting (utilised to transform the data by modifying the weight in the data set). During the training phase, employ regularisation and adversarial training, which are the most common methods for this purpose. Other emerging approaches include decentralised learning, fair linear regression, DeepFair, multimodal models and fairlet clustering. During the post-training phase, implement equalised odds, calibrated equalised odds, and reject option classification.

4.3. Specific recommendations for cities as deployers and regulators

- **Establish local standards for fair AI**. Discrimination and inequalities can manifest differently depending on the context, affecting individuals or social groups in various ways and at different scales. Therefore, cities must implement fairness standards for urban AI that consider these local specificities. These standards should incorporate general principles while integrating local considerations. The goal is to ensure that urban AI does not reinforce existing discrimination or create new forms of bias that disproportionately affect the most vulnerable. These standards should be developed in consultation with local communities and cover the entire AI life cycle.

- **Establish procurement standards for fair AI**. Cities must ensure that entities providing them with services align with fair AI principles. This requires establishing procurement mechanisms that oblige service providers to comply with the city's fair AI standards. Providers must meet compliance requirements regarding their algorithms if they are to be used by the city. For instance, San Jose-led GOV AI in the USA has adopted and introduced the aforementioned AI FactSheet for Third-Party Systems. It is a harmonised template for vendors to provide detailed information about their AI products, covering aspects such as system purpose, training data, model details, performance metrics, bias management, robustness and human-computer interaction.

- **Implement urban laws that ensure the right to justification**. This right allows individuals affected by an AI system to understand the reasoning behind an algorithmic decision, enabling citizens to comprehend and control how they are treated by these systems. When this right is not adequately respected, individuals must have the ability to challenge and modify the underlying parameters of the decision. Therefore, cities must consider, throughout the process, whether to deploy or withdraw an AI system, particularly if an individual's request for explanation cannot be fulfilled. This measure allows individuals facing discrimination to assert their digital rights.

- **Establish advisory bodies to investigate, prevent and mitigate potential malicious uses of AI**. Local governments should set up multidisciplinary advisory bodies that include community organisations, academia, businesses and other stakeholders. These bodies will play an audit role to limit AI-related discrimination. They will assess the city's AI models based on fairness metrics. Their evaluation will (1) identify potential biases that could affect fairness, (2) select metrics to measure the fairness of AI systems, and (3) mitigate the impact caused by these biases. Additionally, they will act as advisory bodies to guide cities in their actions and policies regarding fair AI.

5. Limitations

Achieving fairness in AI is complex. Interventions aimed at achieving fairness in urban AI can create tensions with the very objectives of the algorithms themselves. This implies that cities must adopt a compromise-based approach to balance gains and benefits, while prioritising the well-being of individuals and communities. However, this principle can seem abstract, leaving room for divergent interpretations, which complicates the operationalisation of success and impact measures (Sadek et al., 2024). Therefore, cities need to implement a local approach to operationalising fairness and non-discrimination in urban AI. This holistic approach considers the socioeconomic and cultural configuration of the city throughout the entire AI life cycle.

From a technical perspective, fair urban AI requires diverse human resources and adapted infrastructure (Du et al., 2023; Marvin et al., 2022; Yigitcanlar et al., 2020, 2023). This, in turn, necessitates significant financial investments (Bettoni et al., 2021). Additional costs are also needed for continuous training and education of staff and communities (Sadek et al., 2024; Varanasi, 2023). Cities must also anticipate legal

and compliance costs, including audits and system adjustments to meet regulatory standards. These investments can represent substantial expenses, especially for small and medium-sized cities.

To overcome these limitations, cities can rely on networking. These networks provide opportunities for knowledge sharing, policy innovation and coordinated responses to global issues. Some examples are:

Cities Coalition for Digital Human Rights: a platform to promote an inclusive and democratic development of new technologies in cities.

City AI Connect: A global learning community and digital platform for cities to trial and advance the use of generative artificial intelligence to improve public services.

GovAI: A coalition composed of over 1,000 members and over 350 local, state and federal entities united in the mission to promote responsible and purposeful AI in the public sector.

AI4Cities: A project that enabled Helsinki, Amsterdam, Copenhagen, Greater Paris, Stavanger and Tallinn to challenge the market to come up with AI-based solutions to reduce CO2 emissions in their energy and mobility domains.

References

Barocas, S., Hardt, M., and Narayanan, A. "Classification". In: Fairness and Machine Learning: Limitations and Opportunities. Cambridge: MIT Press, 2023a.

Barocas, S., Hardt, M., and Narayanan, A. "Relative notions of fairness". In: Fairness and Machine Learning: Limitations and Opportunities. Cambridge: MIT Press, 2023b.

Bettoni, A. *et al*. "An AI adoption model for SMEs: A conceptual framework". *IFAC-PapersOnLine*, 2021, 54(1), p. 702–708.

Calzada, I. "The right to have digital rights in smart cities". Sustainability (Switzerland), 2021, 13(20).

Cugurullo, F., *et al*. "Urban AI in China: Social control or hyper-capitalist development in the post-smart city?". *Frontiers in Sustainable Cities*, 2022.

Dauvergne, P. "Facial recognition technology for policing and surveillance in the Global South: A call for bans". *Third World Quarterly*, 2022, 43(9), p. 2325–2335. Routledge.

Du, J. *et al*. "Artificial intelligence-enabled participatory planning: A review". *International Journal of Urban Sciences,* 2023, 28(2), p. 183-210.

Fainstein, S. "Planning and the Just City". In: Marcuse, P., ed. Searching for the Just City: Debates in Urban Theory and Practice. London: *Routledge*, 2009.

Ferrara, E. "Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies". *Sci*, 2023, 6(1), p. 3.

Fincher, R., and Iveson, K. "Justice and injustice in the city". *Geographical Research*, 2012, 50(3), p. 231–241.

Giovanola, B., and Tiribelli, S. "Weapons of moral construction? On the value of fairness in algorithmic decision-making". *Ethics and Information Technology*, 2022, 24(1), p. 3.

González-Sendino, R. *et al*. "A review of bias and fairness in artificial intelligence". *International Journal of Interactive Multimedia and Artificial Intelligence*, 2023.

Hall, P. "Cities of Tomorrow: An Intellectual History of Urban Planning and Design Since 1880". 4th ed. *Wiley-Blackwell,* 2014.

Harvey, D., and Potter, C. "The Right to the Just City". In: Marcuse, P., ed. Searching for the Just City: Debates in Urban Theory and Practice. *Routledge*, 2009.

Holdsworth, J. "What is AI bias?". *IBM*. 2023.

Kleinberg, J. *et al*. "Discrimination in the Age of Algorithms". *Journal of Legal Analysis,* 2018, 10(2005), p. 113–174.

Lefebvre, H. The Right to the City. 1968.

Lyu, Y. *et al*. "IF-City: Intelligible fair city planning to measure, explain, and mitigate inequality". *IEEE Transactions on Visualization and Computer Graphics*, 2023.

Martinez, E., and Kirchner, L. "The secret bias hidden in mortgage-approval algorithms". *The Markup*, August 2021.

Marvin, S. *et al* . "Urban AI in China: Social control or hyper-capitalist development in the post-smart city?". *Frontiers in Sustainable Cities*, 2022, 4, p. 1030318.

O'Neil, C. "Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy". First edition. *Crown*, 2016.

Rezende, I. N. "Facial recognition in police hands: Assessing the 'Clearview case' from a European perspective". *New Journal of European Criminal Law,* 2020, 11(3), p. 375–389.

Sadek, M. *et al*. "Challenges of responsible AI in practice: Scoping review and recommended actions". *AI & SOCIETY*, 2024.

Sanchez, T. W., Brenman, M., and Ye, X. "The ethical concerns of artificial intelligence in urban planning". *Journal of the American Planning Association*, 2024, 0(0).

Sherman, S. "The Polyopticon: A diagram for urban artificial intelligences". *AI and Society*, 2023, 38(3), p. 1209–1222.

UNESCO. "Readiness Assessment Methodology: A Tool of the Recommendation on the Ethics of Artificial Intelligence". UNESCO, 2023.

UN-Habitat. "Mainstreaming Human Rights in the Digital Transformation of Cities: A Guide for Local Governments". United Nations Human Settlements Programme, 2020.

UN-Habitat. "Human Rights in the Digital Era. United Nations Human Settlements Programme", 2023, p. 1–56.

Varanasi, R. A. "'It is Currently Hodgepodge': Examining AI/ML Practitioners' Challenges During Co-production of Responsible AI Values". 2023.

Wachter, S., Mittelstadt, B., and Russell, C. "Why Fairness Cannot Be Automated: Bridging the Gap Between EU Non-Discrimination Law and AI". *Computer Law & Security Review*, 2021, 41, p. 105567.

Weinkauf, D. "Privacy Tech-Know Blog: When Worlds Collide – The Possibilities and Limits of Algorithmic Fairness (Part 1)". *Office of the Privacy Commissioner of Canada,* April 5, 2023.

Yigitcanlar, T., Agdas, D., and Degirmenci, K. "Artificial Intelligence in Local Governments: Perceptions of City Managers on Prospects, Constraints, and Choices". *AI and Society,* 2023, 38(3), p. 1135–1150.

Yigitcanlar, T. *et al*. "Responsible Urban Innovation with Local Government Artificial Intelligence (AI): A Conceptual Framework and Research Agenda". *Journal of Open Innovation: Technology, Market, and Complexity*, 2021, 7(1), p. 1–16.

Yigitcanlar, T. *et al*. "Contributions and Risks of Artificial Intelligence (AI) in Building Smarter Cities: Insights from a Systematic Review of the Literature". *Energies,* 2020, 13(6), 1473.

María Pérez-Ortiz
Associate Professor, Centre for AI, University College London, UK

1. Introduction

By 2050, the United Nations predicts that nearly 70% of the global population will live in urban areas (as opposed to the current 56% (UN-Habitat, 2022)). As the world continues to urbanise at an unprecedented rate, the challenges that cities face – ranging from mobility, provision of services and housing, pollution and urban health, and resource usage – are growing ever more pressing. With this rapid urban growth comes the urgent need to develop innovative solutions to ensure that cities are livable, supportive of human development, efficient and environmentally friendly.

One of the most discussed developments in addressing these challenges is the integration of technology, and more specifically Artificial Intelligence (AI), into urban settings. However, as AI becomes more widely adopted, concerns about its sustainability – both environmental and social – are emerging. In this work, we explore the concept of sustainable AI, focusing on the role it should have in the deployment of these technologies in urban settings. We analyse the environmental, social and economic considerations of AI deployment in cities, highlighting the benefits, challenges and future directions for AI in the quest for sustainable and equitable urban futures.

2. The pursuit of sustainability

Sustainability, and more specifically sustainable development, as it is currently defined by the United Nations (Keeble, 1988), relates to the ability to make development meet present and future human needs (e.g. health and well-being, quality education, decent work, social equality) and how to do so within socioecological limits at present. Sustainability is commonly divided into three pillars:

- **Environmental sustainability**: living within the means of our natural resources and protecting and supporting our ecosystems. In urban settings, a key challenge is reducing greenhouse gas emissions

and improving urban air quality. For example, cities can implement low-emission zones, expand public transportation and promote energy-efficient buildings to reduce urban pollution and mitigate the heat island effect.

Sustainable AI includes the use of AI for social, environmental, and economic sustainability, but also, importantly, the sustainability of AI, addressing a variety of concerns, including its energy consumption, resource use, and social equity.

- **Social sustainability**: persistently achieving good social well-being. In cities, this often means fostering inclusivity, accessibility, equity and more sustainable living in urban developments. A current challenge is creating affordable housing and ensuring equitable access to green spaces, transportation and essential services. For instance, cities can design mixed-use neighbourhoods that are walkable, prioritise social integration and landscape with native plant species to integrate greener environments.

- **Economic sustainability**: using resources efficiently and responsibly. The economic dimension of sustainability focuses on fair, green and circular economies rather than simply continuous economic growth. In the context of urban sustainability, the focus is on developing resilient, low-carbon economies. One example is promoting sustainable urban infrastructure that supports green jobs and circular economy principles, such as creating sustainable urban mobility systems (e.g. bike-sharing, electric vehicle infrastructure) and supporting local green industries.

While we often think of sustainability as just the environmental pillar, the three examples of social, environmental and economic sustainability above demonstrate the interconnectedness of all the pillars and the need to consider them together. As environmental philosopher John Muir put it: "When we try to pick out anything by itself, we find it hitched to everything else in the Universe" (Muir, 1911). Similarly, cities are complex systems, requiring considerations of the system as a whole and approaches that multi-solve goals and take into account the social, environmental and economic pillars.

3. Technology meets sustainability

Could technological advances influence our ability to secure a sustainable future? (GACGC, 2019). It is crucial to understand the impact of technology on sustainability, especially as disruptive technologies such as AI accelerate global and human-led change at an unprecedented pace and without a clear unified agenda.

In the last few years, we have seen a new set of desired principles towards sustainability emerge for AI systems (Vinuesa, 2020; Van Wynsberghe, 2021). In particular, sustainable AI refers to a rapidly evolving framework that aims to shape the development, deployment and use of AI technologies in a way that is environmentally, socially and ethically responsible, seeking to balance the tensions between the risks and opportunities of AI.

At its core, sustainable AI also considers the three major dimensions of sustainability: social, environmental and economic. However, sustainable AI is more than the sum of its parts. Sustainable AI includes the use of AI for social, environmental and economic

sustainability, but also, importantly, the sustainability of AI (Van Wynsberghe, 2021), addressing a variety of concerns, including its energy consumption, resource use and social equity, to ensure that AI solutions contribute positively to society and the environment over time.

One of the core challenges in developing sustainable AI is addressing the environmental footprint of AI itself. AI systems, particularly large-scale models like those used in deep learning, require vast computational power, resulting in significant energy consumption and carbon emissions. For example, in 2019 it was estimated that training a single large language model can emit as much carbon as five cars over their lifetimes (Strubell et al., 2019), an estimate that is likely to have increased significantly since then. This is only training, not considering the footprint of its usage (e.g. it is estimated that making an image with generative AI uses as much energy as charging your phone (Heikkilä, 2023)). In addition to operational emissions, it is important to consider embodied emissions – the carbon footprint associated with the production, transportation and disposal of the hardware used for AI, such as servers and GPUs. Sustainable AI seeks to minimise both operational and embodied environmental impacts by promoting more energy-efficient algorithms such as model distillation or quantisation.

Sustainable AI is not just about a subset of technologies designed specifically for sustainability, but about reshaping the entire field of AI to ensure it consistently contributes to long-term social, environmental and economic well-being.

Beyond its environmental impacts, social sustainability in AI is also a pressing concern. AI systems, if not carefully designed, can reinforce existing social inequalities, through biased algorithms or unequal access to AI technologies. Sustainable AI calls for the creation of AI systems that promote social inclusion and fairness, ensuring that marginalised groups are not harmed by AI-driven decisions in areas such as hiring, housing or criminal justice. This involves embedding ethical considerations and human rights into the design and implementation of AI, with robust transparency, accountability and governance frameworks to guard against unintended harms. Further exploration of fairness in AI systems is discussed in the fairness chapter of this monograph.

In terms of economic sustainability, AI has the potential to both drive and disrupt sustainable development. AI can optimise urban infrastructure, energy grids and transportation systems, promoting more efficient resource use and reducing waste in cities. It can also enable the transition to a circular economy by improving processes like supply chain management, waste reduction and product life cycle optimisation. However, if not carefully managed, AI may exacerbate economic inequality, automating jobs without creating new employment opportunities or increasing wealth concentration. Sustainable AI, therefore, advocates for AI-driven economic models that prioritise long-term societal benefits over short-term profit, ensuring that the economic gains from AI are distributed equitably.

Some of the questions proposed for inclusion in the design of AI systems by frameworks of sustainable AI include (Vinuesa, 2020; Van Wynsberghe, 2021):

- What are the trade-offs between the direct and indirect impacts of AI technology on society, the environment and the economy? How can we design AI systems to be more sustainable from the outset?

What risk assessment frameworks can help us anticipate unintended consequences before they arise?

- How can we address the broader sociotechnical system surrounding AI, including the social impacts on individuals who use or are affected by these technologies? What steps can we take to develop AI that aligns with the preservation of environmental resources for current and future generations, supports sustainable economic models and respects the core societal values of different communities?

- How can we promote change throughout the entire AI life cycle – from idea generation, training and fine-tuning to evaluation, implementation and governance – towards greater ecological sustainability and social equity? What measures are necessary to ensure AI systems operate within the planet's ecological limits, such as energy consumption, freshwater use and reliance on scarce minerals?

AI technologies in cities have a significant environmental impact due to their reliance on data centres and computational resources. A sustainable approach to urban AI would involve optimising these systems for energy efficiency.

Ultimately, sustainable AI advocates for an approach where AI serves as a force multiplier for sustainability goals, enhancing efforts to mitigate climate change, reduce inequality and foster inclusive, resilient economies, while also ensuring that the development and use of AI technologies themselves are aligned with principles of sustainability. It is important to recognise that technologies, including AI, can pose both extrinsic and intrinsic risks to sustainability. In the case of intrinsic risks, even when a technology is not directly applied to sustainability challenges, if it is faulty, non-robust or unfair, it may unintentionally undermine sustainability goals by exacerbating inequality, environmental harm or economic instability through indirect channels. Therefore, sustainable AI is not just about a subset of technologies designed specifically for sustainability, but about reshaping the entire field of AI to ensure it consistently contributes to long-term social, environmental and economic well-being.

4. Sustainable AI in urban settings

While more frameworks for the sustainability of AI are starting to emerge (Vinuesa, 2020; Van Wynsberghe, 2021; Wu, 2022; Wilson, 2022; Nishant, 2020), very few works at present focus on urban futures (Yigitcanlar, 2020; Bibri, 2021; Pastor-Escuredo, 2022). Towards this goal, we aim to introduce a perspective that exemplifies the foundations that would be needed to ensure that AI systems deployed in cities are not only technologically advanced but also responsible, equitable and beneficial for both the environment and urban populations.

Environmental impact in urban AI systems. One of the relevant challenges of AI is its environmental impact, which includes but is not restricted to its significant energy consumption and freshwater usage (Luccioni, 2024). In urban settings, where AI is increasingly used in applications such as smart traffic systems, energy grids and building management, the cumulative demands of these systems can become substantial. AI technologies in cities have a significant environmental impact due to their reliance on data centres and computational resources. A sustainable approach to urban AI would involve optimising these systems for energy efficiency through methods such as tiny

machine learning, green computing software engineering practices, knowledge distillation, model pruning or quantisation. This could also mean learning from smaller, high-quality datasets (i.e. doing more with less), using renewable energy, supporting sustainable consumption and production patterns, and minimising the carbon footprint of city-wide AI deployments. For example, AI-powered smart grids could dynamically adjust energy usage based on real-time data, reducing waste and supporting the integration of renewable energy sources like solar and wind. Federated learning is another promising approach for sustainability in urban AI. Rather than relying on centralised data centres for large-scale model training, federated learning enables smaller models to be trained directly on decentralised devices, such as Internet of Things (IOT) sensors embedded in urban infrastructure. This reduces the need to transmit vast amounts of data to centralised servers, cutting down on energy-intensive data processing and storage. By leveraging existing local computing resources, federated learning also reduces the overall demand for new hardware and mitigates the environmental footprint of large-scale AI operations. Additionally, it enhances privacy and data security by keeping sensitive information on local devices, reducing the need for data sharing while supporting sustainable AI practices.

The true social sustainability of these systems must also consider the ethics of their development process, ensuring fair practices across the entire AI life cycle. This helps advance social sustainability by fostering more just and inclusive cities while addressing global inequities in AI production.

Social sustainability: equity and fairness. As cities increasingly adopt AI to power services such as security, healthcare and public resource allocation, it is essential that these systems contribute to social sustainability by promoting fairness, equity and inclusivity. In urban planning, for example, sustainable AI could be leveraged to identify and address inequalities, such as ensuring underserved neighbourhoods receive equitable access to transportation, healthcare and education. However, the social component of sustainable AI not only involves the purpose for its use, designing algorithms that minimise bias and ensuring marginalised communities are not harmed by AI-driven decisions, but also addressing the ethical implications of how AI systems are developed and deployed. Many AI systems are trained and maintained by underpaid and overworked workers in Global South communities (Rowe, 2023), who are often employed by third-party companies. This labour, crucial to training many AI systems, highlights deep inequalities in the global AI supply chain, as these workers often face poor working conditions while bearing the toll of repetitive, underappreciated tasks. The true social sustainability of these systems must also consider the ethics of their development process, ensuring fair practices across the entire AI life cycle. This helps advance social sustainability by fostering more just and inclusive cities while addressing global inequities in AI production.

Ethical governance and accountability. Urban AI systems must be governed by strong ethical frameworks that prioritise transparency and accountability. City governments and stakeholders should ensure that AI systems are explainable and that decision-making processes are clear to the public. This would build trust and ensure that any errors or unintended consequences can be identified and addressed promptly. For example, AI systems used for surveillance or law enforcement in cities should be designed with clear accountability structures, protecting citizens' privacy and civil rights. Further exploration on the operationalisation of these principles in AI systems is discussed in the transparency and accountability chapter of this monograph.

For AI to be sustainable in urban environments, it must also be economically viable in the long run. This involves developing AI systems that integrate seamlessly with existing city infrastructure, scale to meet future demands and are built for long-term use.

Economic sustainability in cities. For AI to be sustainable in urban environments, it must also be economically viable in the long run. This involves developing AI systems that integrate seamlessly with existing city infrastructure, scale to meet future demands and are built for long-term use. Cities can support circular economy models by encouraging the reuse and recycling of AI technologies, data and hardware, thereby reducing waste and lowering costs.

Alignment with urban sustainability goals. AI systems deployed in cities should not only support urban sustainability goals – such as reducing pollution, enhancing public health and improving quality of life – but also ensure that the use of AI technologies themselves contributes to sustainability. One way this can be achieved is by repurposing the energy and resources used by AI infrastructure. For example, reusing excess heat from data centres – a significant byproduct of AI's computational demands – can contribute to urban sustainability by reducing overall energy consumption. In Stockholm, the Stockholm Data Parks project[1] has shown how waste heat from data centres can be redirected to heat residential and commercial buildings, demonstrating how AI infrastructure can be integrated into a circular economy model, aligning with climate goals while reducing public energy needs. Beyond resource efficiency (connected with the concept of the sustainability of AI), AI can optimise urban systems for sustainability. By leveraging AI to enhance resource management, reduce energy consumption and support climate resilience initiatives, cities can address pressing challenges like climate change and urbanisation. For instance, urban areas are particularly vulnerable to the impacts of climate change, such as rising temperatures and extreme weather events. AI can play a crucial role in increasing the climate resilience of cities by providing advanced predictive analytics and early warning systems for climate-related risks. These systems can alert authorities to potential environmental hazards and enable rapid response, helping to mitigate the health and environmental impacts of urban pollution.

AI for urban planning and development. AI is transforming the way cities are planned and developed (Jha, 2021) fostering greater social, environmental and economic sustainability. By analysing large datasets on population growth, land use, transportation patterns and environmental factors, among many others, AI can help urban planners and policymakers design more sustainable and efficient cities using so-called digital twins. For example, AI models can predict how changes in infrastructure, such as the construction of new roads or public transit systems, will impact traffic patterns, pollution levels and energy consumption. This helps planners anticipate the future and engage in responsible foresight, allowing for more informed decisions that promote long-term sustainability. AI can also be used to optimise land use and zoning policies, ensuring that urban development is balanced with the preservation of green spaces and natural resources. This is particularly important in rapidly growing cities, where the demand for housing and infrastructure often leads to urban sprawl and the loss of valuable ecosystems. However, beyond its application to urban systems, it is crucial that the development and deployment of AI itself aligns with sustainable practices. While digital twins are powerful tools for simulating urban planning scenarios, their sustainability depends

1. https://stockholmdataparks.com/

on the efficiency of the underlying AI models and the infrastructure supporting them. It is worth noting however, as reported by many studies (Andersson, 2021), that digital twins built with AI can be a more resource efficient approach than their physics-based counterpart simulations, running on laptop CPUs in seconds as opposed to necessitating supercomputers for days.

5. Policy recommendations and concluding remarks

Innovation and technology will play an increasingly central role in planning for sustainable urban futures (UN-Habitat, 2022). As we discuss next with our list of policy recommendations, the design and deployment of technology should be tailored to suit the large diversity of the urban context:

a. Environmental sustainability recommendations

- The urgency to decarbonise urban economies should drive the **convergence of green and smart technologies**. Policies should emphasise energy efficiency, environmental preservation and resilience. This includes the establishment of green AI standards that prioritise energy-efficient algorithms and hardware, as well as creating circular economies surrounding data centres, e.g. recycling excess heat. Life cycle management policies should promote the responsible sourcing, reuse and recycling of AI hardware to minimise electronic waste.

- Impact assessments should carefully weigh whether deploying **AI for sustainability projects justifies the environmental cost of the technology**, as highlighted by previous work (Dixon, 2022). New frameworks are essential to measure and compare the full life cycle costs of AI, ensuring a comprehensive evaluation of its sustainability.

- **AI energy star rating frameworks** are beginning to emerge (Luccioni, 2024) and these should be added to IOT urban devices, offering users valuable insights to better understand the environmental impact of the tools they use and to adopt them more responsibly.

- Public-private partnerships and collaboration can drive the development of sustainable AI technologies in urban areas. Creating **AI for sustainable cities consortiums** can foster partnerships between governments, tech companies and research institutions to tackle urban challenges such as energy management, transportation and waste reduction. Cities should incentivise sustainable AI development by offering tax credits or subsidies to companies developing environmentally friendly AI solutions.

- Urban resilience and smart infrastructure should be supported through **policies that encourage AI for climate and biodiversity resilience**. This includes the use of AI-driven early warning systems for natural disasters, tipping points of biodiversity loss and extreme climate events.

b. Social sustainability recommendations

- Since all the dimensions of sustainable AI are intertwined, ethical and responsible AI deployment are also a critical dimension in urban environments. AI systems must be audited for fairness to prevent discrimination and social inequality. The creation of **local AI ethics boards** should ensure that urban AI projects adhere to privacy, fairness and accountability standards.

- Data privacy and security are also key areas of focus. **Strong urban data privacy laws** should be enacted to protect personal data collected from sensors, cameras and mobile apps. This includes anonymisation and the use of explicit consent. In addition, secure and transparent frameworks for data sharing between governments, private companies and AI developers are necessary to ensure responsible use of citizen data without compromising privacy.

- To foster public support and understanding, policies should promote **public engagement and digital literacy**. Cities should encourage **participatory governance models** that involve citizens in AI decision-making processes, while also launching digital literacy campaigns to educate the public on AI technologies, their impacts and how to protect their rights.

- Ensuring **equal access to AI-driven public services**, particularly for marginalised and underserved communities, is essential to promote inclusivity.

- Fostering **open science** is essential, enabling public audits of these systems while ensuring robust cybersecurity measures are in place to protect sensitive data and utilities. Transparency is also key, with regulations requiring AI systems used in public services to be explainable, enabling both stakeholders and the public to understand how decisions are made.

c. General recommendations

- To address the economic impact of AI, policies should **support job transition and workforce development**. Public funding should be allocated to reskilling programmes that help workers transition into new jobs, particularly in emerging sectors where automation may cause job displacement. Promoting the growth of AI-based green jobs, such as those related to renewable energy management and sustainable urban infrastructure, can further drive sustainable economic growth.

- Monitoring and accountability frameworks are essential to ensure AI systems align with sustainability goals over time. **Mandatory AI impact assessments**, similar to environmental impact assessments, should evaluate the social, economic and environmental effects of AI deployment in cities. Continuous **monitoring and auditing** of urban AI systems can help ensure they remain adaptable to new challenges and ethical considerations.

- **Include considerations into environmental, social and governance standards** that account for the sustainability of the data, algorithms and computational resources used by businesses, as well as the support provided to renewable energy sources and circular computational economies.

- Regulatory standards for smart city infrastructure must ensure that **AI technologies are adaptable, interoperable and scalable for future urban needs**, especially in areas like traffic management, waste reduction and energy efficiency.

- Finally, at the global level, **international collaboration and standardisation** should be encouraged. Cities should work together to develop global sustainability standards for AI and share best practices (Strubell, 2019), ensuring alignment with international goals. Platforms for cross-city knowledge sharing can help accelerate the adoption of sustainable AI practices worldwide.

References

Andersson, T. R., et al. "Seasonal Arctic sea ice forecasting with probabilistic deep learning". *Nature Communications* 12.1 (2021): 5124.

Bibri, S. E. "Data-driven smart sustainable cities of the future: Urban computing and intelligence for strategic, short-term, and joined-up planning". *Computational Urban Science* 1.1 (2021): 8.

Dixon, B., Pérez-Ortiz, M. and Bieker, J. "Comparing the carbon costs and benefits of low-resource solar nowcasting". NeurIPS Workshop on Tackling Climate Change with Machine Learning (2022).

German Advisory Council on Global Change, *Towards Our Common Digital Future*, Flagship Report, 2019.

Heikkilä, M. "Making an image with generative AI uses as much energy as charging your phone", *MIT Technology Review* (2023).

Jha, A. K. *et al*. "A review of AI for urban planning: Towards building sustainable smart cities". *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2021.

Keeble, B. R. "The Brundtland Report: 'Our Common Future'". *Medicine and War*, vol. 4, no. 1 (1988), p. 17-25.

Luccioni, S., Trevelin, B. and Mitchell, M. "The Environmental Impacts of AI—Primer", 2024.

Muir, J. "My first summer in the Sierra", in: *British Politics and the Environment in the Long Nineteenth Century*. Routledge, 1911 (republished in 2023), p. 291-296.

Nishant, R., Kennedy, M. and Corbett, J. "Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda". *International Journal of Information Management* 53 (2020): 102104.

Pastor-Escuredo, D., Treleaven, P. and Vinuesa, R. "An Ethical framework for artificial intelligence and sustainable cities". *Ai* 3.4 (2022), p. 961-974.

Rowe, N. "Millions of Workers Are Training AI Models for Pennies", *WIRED* (2023).

Strubell, E., Ganesh, A. and McCallum, A. "Energy and policy considerations for modern deep learning research". *Proceedings of the AAAI Conference on Artificial Intelligence.* Vol. 34. No. 09. 2020.

UN Habitat, "Envisaging the Future of Cities", *World Cities Report 2022*.

Van Wynsberghe, A. "Sustainable AI: AI for sustainability and the sustainability of AI". *AI and Ethics* 1.3 (2021), p. 213-218.

Vinuesa, R. *et al*. "The role of artificial intelligence in achieving the Sustainable Development Goals". *Nature Communications* 11.1 (2020), p. 1-10.

Wilson, C. and Van Der Velden, M. "Sustainable AI: An integrated model to guide public sector decision-making". *Technology in Society* 68 (2022): 101926.

Wu, C-J. *et al.* "Sustainable AI: Environmental implications, challenges and opportunities". *Proceedings of Machine Learning and Systems* 4 (2022), p. 795-813.

Yigitcanlar, T. and Cugurullo, F. "The sustainability of artificial intelligence: An urbanistic viewpoint from the lens of smart and sustainable cities". *Sustainability* 12.20 (2020): 8548.

PART II. CASE STUDIES OF URBAN AI GOVERNANCE FRAMEWORKS

- CASE STUDY 1: BARCELONA
- CASE STUDY 2: AMSTERDAM
- CASE STUDY 3: NEW YORK
- CASE STUDY 4: SAN JOSÉ
- CASE STUDY 5: DUBAI
- CASE STUDY 6: SINGAPORE

Alexandra Vidal D'oleo
Research Fellow and Project Manager, Global Cities Programme, CIDOB

The first report of GOUAI's Atlas of Urban AI (Galceran-Vercher and Vidal, 2024) reveals that while many cities are currently experimenting with artificial intelligence (AI), only a small percentage have implemented policies or overarching strategies to regulate its use and ensure alignment with key ethical principles. The emphasis recently has fallen on tackling immediate urban challenges in a solution-oriented pragmatism. As a result, there is a significant gap between the adoption of AI and the establishment of effective governance frameworks. However, driven by the public discourse and a rising tide of opinion pushing for global regulation of algorithms and AI, some local governments have taken the lead in creating their own governance frameworks. This trend is expected to grow exponentially in the years to come.

The following pages offer a collection of case studies from cities worldwide that have strived to establish AI local governance frameworks by adopting different policy mechanism to govern AI comprehensively. All policy mechanisms fall under one of the following categories: (1) principles, strategies and guidelines; (2) local regulations and laws; (3) transparency and explainability mechanisms; (4) algorithmic impact assessment; (5) audits and regulatory inspection; (6) human oversight, accountability, hearing and appeal procedures; (7) procurement conditions; (8) external/independent oversight and advisory bodies; (9) alliances, communities of practice and learning groups; (10) capacity-building programmes; (11) promotion of local innovation, knowledge and experimentation; (12) community engagement; (13) data governance; and (14) other policies and measures. It is worth noting that although some of the policies presented are believed to be under development, they may have already been implemented without notice.

The chosen cities vary in geographic location, size and income per capita. The following case studies are presented:

1.Barcelona (Spain)
2.Amsterdam (The Netherlands)
3.New York (United States of America)
4.San José (United States of America)
5.Dubai (United Arab Emirates)
6.Singapore (Republic of Singapore)

Case study 1: Barcelona

Population: 1,655,956 (2023)	**Income per capita:** €31,531 (2022)	**Region:** Europe

AI GOVERNANCE FRAMEWORK		
[1] Principles, strategies and guidelines	AI strategy	Municipal strategy on algorithms and data to ethically drive artificial intelligence: defines a set of guiding principles and 20 actions for ethical AI deployment, including: • the use of AI for automated recommendation systems, rather than decision-making systems. • transparency and auditability: the algorithmic models and databases should be accessible, understandable and auditable by the general public. • the establishment of liability regimes for any harm or loss that may occur.
	Local AI principles	Included in the AI municipal strategy: (1) action and human supervision; (2) technical robustness and security; (3) privacy and data governance; (4) transparency; (5) diversity, equity and inclusion ; (6) social and environmental commitment; (7) responsibility, democratic control and accountability
	Internal protocols for AI use	Definition of work methodologies and protocols for implementing algorithmic systems: defines the mechanisms for each stage of the tendering and implementation of AI systems by the city council and establishes the governance and supervision bodies that will ensure alignment with ethical principles.
[3] Transparency and explainability mechanisms	Public algorithm register (*)	Creation of a municipal register of current and future algorithms that impact municipal procedures and services. The register will be public and will also serve to classify algorithms according to the risk they pose, with clear explanations for citizens and other interested parties. For each registered algorithm, it will incorporate a public contact point that citizens can contact.
	Algorithmic transparency standard	Development of a common algorithmic model that ensures the appropriate use of data. Project developed with 7 other European cities within the framework of Eurocities.
	Municipal website disclosing all AI relevant information	Barcelona Digital City: municipal website disclosing all resources available in the city to boost digitalisation, including all relevant information regarding AI projects and initiatives.
[4] Algorithmic impact assessment	Risk assessment and management	Risk analysis (included in the protocol): The AI Technical Office evaluates the algorithms in use by the municipality and issues a report including the risk assessment. Depending on the risk determined by the office, the following steps are taken. All algorithms categorised under "unacceptable risk" are rejected, whilst "high risk" algorithms undergo a mandatory algorithmic impact assessment by the municipal Transversal Commission (see below).
	Human rights impact assessment	Mandatory algorithmic impact assessments for high-risk systems (included in the strategy and protocol).
[5] Audits and regulatory inspection	Audits	Mandatory audits for high-risk systems (included in the strategy and protocol). The conclusions will be made public through the algorithm register.

[7] Procurement conditions	Procurement clauses*	Inclusion of clauses related to digital rights in the tendering of solutions based on artificial intelligence.
[8] External/ independent oversight and advisory bodies	Advisory council	Advisory Council on AI, Ethics and Digital Rights: comprised of 15 independent and multidisciplinary experts. Its mission includes advising the government in the use of AI, conducting algorithmic impact studies on high-risk algorithmic systems and assessing the development of the municipal AI strategy.
[9] Alliances, communities of practice and learning groups	Community of practice member	Cities Coalition for Digital Rights (CC4DR) (founding member): a global city network of 60+ members working in the greenfield of digital rights-based policymaking.
[10] Capacity-building programmes	Municipal capacity-building	Municipal staff training (included in the strategy).
	Municipal body	Transversal commission to encourage ethical AI: comprising 25 members, its mission is to guide and align municipal policies to develop tools that use AI and promote interdepartmental collaboration. They developed the AI strategy.
[11] Promotion of local innovation, knowledge and experi-mentation	Local AI observatories	Global Observatory for Urban AI (GOUAI): an initiative of Barcelona, led by CIDOB and established in collaboration with Amsterdam and London, within the framework of the CC4DR, and the support of UN-Habitat. It conducts research on urban AI from an ethical standpoint.
[12] Community engagement	Public engagement	Communication channels with the public (included in the strategy). Promotion of spaces for reflection and debate on the impact of AI on public services (included in the strategy).
[13] Data governance	Data transparency	Open data BCN: the city's open data portal. It aims to maximise available public resources, making the data generated or stored by public bodies accessible, free and usable to all.
	Data rights	Decode project: a collaborative EU initiative to strengthen citizens' data rights and put them in control of their data, as well as enable them to share it for the common good.

* Planned policy mechanism not fully implemented in December 2024.

Case study 2: Amsterdam

Population: 921,402 (2022)	**Income per capita:** €54,700 (2022)	**Region:** Europe

AI GOVERNANCE FRAMEWORK		
[1] Principles, strategies and guidelines	AI agenda	Amsterdam Intelligence Agenda (2020-2024): highlighted the city's goals in the area of algorithms, particularly AI. The objectives focused on taking action that improved the quality of life for Amsterdam's residents, reduced the harmful effects of digitalisation and boosted its beneficial outcomes. Amsterdam's vision on AI (2024): in its new policy, the city outlines how AI should be integrated into urban life and how it should influence the city. The vision was developed through discussions with Amsterdam residents, experts and municipal staff. To put it into operation, the municipality is currently developing a new AI Agenda, expected to be published in 2025. It will provide guidelines for the responsible use of AI within the municipality, ensuring the technology is applied in an ethical, inclusive and sustainable way.
	Local AI principles	Tada principles endorsement (from the Tada manifesto) which guide the Amsterdam Digital Agenda and also apply to the AI field: (1) inclusive; (2) control; (3) tailored to the people; (4) legitimate and monitored; (5) open and transparent; (6) from everyone – for everyone. Local AI guiding principles (as included in Amsterdam's vision on AI (2024): (1) human-centric; (2) reliable; (3) future-proof.
	Guidelines, playbooks and manuals	Algorithms playbook: guideline document that sets out the city's integral approach and policy tools for the responsible use of algorithms. The Algorithm Lifecycle Approach consists of seven tools to manage, assess risks and investigate algorithms throughout their life cycle, namely: (1) algorithm register; (2) contractual terms; (3) objections procedure; (4) governance definition and life cycle model; (5) audit; (6) bias analysis model; and (7) human rights impact analysis model.
		The Fairness Handbook: standard for bias analysis, a step-by-step plan to evaluate a model for biases and mitigate their effects.
	Internal protocols for AI use	Governance establishment and life cycle model: specifies tasks and responsibilities, measures to be taken to prevent risks when applying algorithms, information to be documented and who is responsible if an algorithm does not meet its intended purpose. The Algorithm Lifecycle Approach describes the process of an algorithm from start to finish. Included in the playbook.
[3] Transparency and explainability mechanisms	Public algorithm register	Algorithm Register: overview of the AI systems and algorithms used by the municipality.
	Algorithmic transparency standard	Development of a common algorithmic model developed by 7 European cities within the framework of Eurocities based on Amsterdam's and Helsinki's transparency standard.
	Municipal website disclosing all AI relevant information	Municipal accessible portal disclosing all AI relevant information and resources, such as landmark projects and strategic documents.

[4] Algorithmic impact assessment	Human rights impact assessment	Human Rights Impact Assessment Model: based on practical lessons learned and existing tools (included in the playbook).
	Bias analysis	Step-by-step plan to evaluate a model for biases, including the following components: (a) defining the ("sensitive") groups to be studied; (b) drafting hypotheses on features that may lead to indirect biases; (c) selecting metrics that fit the project; (d) analysis of direct and indirect bias; (e) analysing bias on non-measurable variables; (f) weighing and reviewing biases found with responsible management; (g) mitigating biases where necessary; and (h) drafting conclusions. The above-mentioned Fairness Handbook provides the guidelines to conduct bias analysis.
[5] Audits and regulatory inspection	Audits	Annual audits: commissioned by the CIO's office and carried out by the Audit Service ACAM. A framework of standards has been developed for these audits (included in the playbook).
[6] Human oversight, accountability, hearing and appeal procedures	Feedback and objections procedures accessible for citizens	Objections procedures and guidelines for objections handlers (included in the playbook).
[7] Procurement conditions	Procurement clauses	Standard Clauses for Procurement of Trustworthy Algorithmic Systems: the pioneer standard stipulating the contractual conditions and information requirements needed from suppliers of procured AI systems.
[8] External/ independent oversight and advisory bodies	External advisory and oversight body	Amsterdam Personal Data Committee: advises the municipality on algorithms, data ethics, digital human rights and the exposure of personal data. Includes ethical assessment in the use of algorithms. The committee upholds transparency by organising public meetings and by issuing opinions.
[9] Alliances, communities of practice and learning groups	Community of practice member	Cities Coalition for Digital Rights (CC4DR) (founding member)
	Multistakeholder AI collaborations	Amsterdam AI: collaboration between the municipality, Amsterdam knowledge institutions, research centres, medical centres and the Amsterdam Economic Board. This collaboration focuses on responsible AI with a human-centred approach. NL AI Coalition: the city is part of the national working group on AI in the public sector, together with Amsterdam AI coalition. Smart Health Amsterdam network: the regional network for data- and AI-driven innovation in the life sciences and health sector in Amsterdam (also part of the Amsterdam AI coalition).
[10] Capacity-building programmes	Municipal capacity-building	Municipal staff training: all officials must take the National AI course. Legal, procurement and auditing services must take regular refresher courses. Municipal AI team creation.

[11] Promotion of local innovation, knowledge and experimentation	Innovative AI centres, hubs and laboratories	Civic AI Lab: set up by the municipality, the public university and a national ministry, the aim is to investigate how AI can counteract social inequality or prevent AI from reinforcing it. DataLab: creates open accountable tech solutions.
[13] Data governance	Data transparency	Data Amsterdam (beta version): the city's open data portal. It aims to maximise available public resources, making the data generated or stored by public bodies accessible, free and usable to all.
	Data rights	Decode project
	Data sharing mechanisms	Amsterdam Data Exchange (AMdEX): for data sharing between organisations through the creation of a digital notary.

Case study 3: New York

Population: 8,258,000 (2023)	**Income per capita:** €44,537 (2022)	**Region:** North America

AI GOVERNANCE FRAMEWORK		
[1] Principles, strategies and guidelines	AI strategy	AI strategy (2021-2023): foundational effort to foster a healthy cross-sector AI local ecosystem. The document established a baseline of information about AI to help ensure decision-makers were working from a shared understanding of the technology and the issues it presented. It included a set of ethics, governance and policy framework.
	AI action plan	AI Action Plan (2023-2025): includes 37 actions across seven initiatives to create governance for the city's use of AI. Mandates an annual progress report of the plan.
	Local AI principles	New York City AI Principles (included in its AI Action Plan): (1) validity and reliability; (2) social responsibility; (3) information privacy; (4) cybersecurity; (5) trust and transparency. Principles included in its former AI Strategy: (1) accountability; (2) fairness; (3) privacy and security; (4) community engagement and participation.
	Guidelines, playbooks and manuals	Guideline on AI Principles and Definitions: specifies the city's AI principles, provides concrete definitions of AI-related terms, and specifies related laws, policies, requirements and processes that apply. Preliminary Use Guidance: Generative Artificial Intelligence: includes terms and definitions, roles and responsibilities, guidance for GenAI use by the municipality, and specifies related laws, policies, requirements and processes that apply.
	Internal protocols for AI use	Internal protocols specified in the Action Plan, including a mandatory annual report by city agencies to expand public AI reporting.
[2] Local regulations and laws	Regulation of controversial AI application	Biometric data protection law for businesses. Recruitment technology required to audit for bias.
[3] Transparency and explainability mechanisms	Municipal website disclosing all AI relevant information	Municipal accessible portal disclosing all AI relevant information and resources, such as landmark projects and strategic documents.
	Municipal directory of procured AI tools (for internal use)	Establishment of an internal directory of procured AI tools and guidance on their appropriate use, shared across agencies to support visibility and access. Included in the action plan.
[4] Algorithmic impact assessment	Risk assessment and management (*)	AI Risk Assessment and Project Review Process (included in the action plan).
[7] Procurement conditions	Procurement clauses (*)	AI-specific procurement standards.

[8] External/ independent oversight and advisory bodies	External advisory and oversight body	AI Advisory Network: brings together independent experts from private industry, academia, labour and civic organisations to support the city's AI efforts on a consultative basis.
[9] Alliances, communities of practice and learning groups	Community of practice member	Cities Coalition for Digital Rights (CC4DR) (founding member).
[10] Capacity-building programmes	Municipal body	NYC Automated Decision Systems Task Force: established and tasked with issuing recommendations addressing how the city ought to manage the use of algorithms. It was the first of its kind in the country and culminated in the publication of an accessible report. Citywide AI Steering Committee: brings stakeholders from across city government together to provide input and oversight AI activities.
[12] Community engagement	Public engagement	Public listening sessions.
[13] Data governance	Data transparency	NYC Open Data: the city's open data portal. It aims to maximise available public resources, making the data generated or stored by public bodies accessible, free and usable to all.

* Planned policy mechanism not fully implemented in December 2024.

Case study 4: San José

Population: 969,655 (2023)	**Income per capita:** €139,761 (2023)	**Region:** North America

AI GOVERNANCE FRAMEWORK		
[1] Principles, strategies and guidelines	AI handbook	City's AI Handbook: provides comprehensive guidance on how to comply with the city's AI policy. It includes: (1) AI policy, (2) AI review (required for all procurements and data initiatives); (3) AI governance (the framework for managing and monitoring the AI life cycle); (4) GenAI guidelines.
	Guidelines, playbooks and manuals	Generative AI Guidelines: was the first step in a collaborative process to develop the city's overall AI policy. Registered users were invited to join the IT Department in a working group to share their experience and co-develop the city's AI policies.
	Local AI principles	San José AI Principles: (1) effectiveness (reliability); (2) transparency; (3) equity; (4) accountability; (5) human-centred design; (6) privacy; (7) security and safety; (8) workforce empowerment
[3] Transparency and explainability mechanisms	Public algorithm register	AI Inventory (Step 5 of the Review process): overview of the AI systems and algorithms used by the municipality.
	Municipal website disclosing all AI relevant information	Municipal accessible portal disclosing all AI relevant information and resources, such as landmark projects and strategic documents.
[4] Algorithmic impact assessment	Risk assessment and management	AI Risk Threshold Analysis model: conducted by the Digital Privacy Office (Step 2 of the Review process).
	Human rights impact assessment	Algorithmic impact assessment: conducted by the municipality when procured AI systems are categorised as medium-high risk by the risk analysis (Step 3 of the Review process).
[6] Human oversight, accountability, hearing and appeal procedures	Feedback and objections procedures accessible for citizens	Public Comment Form: citizens can submit comments on projects that involve a new usage of personal information using the form. Information on new projects can be found online.
	Internal monitoring and reporting	Annual Usage Report: the business-owning department of the AI system must submit an Annual Usage Report detailing: 1. Project summary 2. Required performance metrics 3. Future plans for the technology initiative The public can comment online on data usage and annual updates (Step 6 of the Review process).
[7] Procurement conditions	Internal protocols for AI procurement	AI Review framework: to assess the benefits and risks in municipal procurement. Review process: 1. Procurement request 2. Risk analysis 3. Algorithmic impact assessment (for medium-high risk systems): includes the municipal algorithmic impact assessment and vendor AI factsheet 4. Final review 5. Pre-launch preparation: data usage protocol, training users and AI inventory posting 6. Ongoing monitoring
	Procurement clauses	Vendor AI Factsheet: includes a factsheet and an algorithmic impact assessment questionnaire for the vendor (Step 3 of the review process).

[8] External/ independent oversight and advisory bodies	Multistakeholder advisory and oversight body	AI Advisory Group: led by the municipality, external stakeholders advise city departments and the CIO on the policies and activities related to AI governance. Consists of AI experts from industry, academia, civil rights and members of the public. The Advisory Group meets quarterly, and the decision-making power remains within the municipality.
[9] Alliances, communities of practice and learning groups	Multistakeholder alliance	GovAI Coalition: the San José led coalition brings together public agencies, civil society, academic institutions and companies to promote responsible AI in the public sector. Composed of 1,500+ members and 500+ local, state and federal agencies.
	Community of practice member	Cities Coalition for Digital Rights (CC4DR)
[10] Capacity-building programmes	Municipal working group	AI Working Group (AIWG): employees from various municipal departments discuss AI-related issues and projects in the city. Composed of department AI-leads and potentially other department representatives.
[12] Community engagement	Public engagement	If a procured AI system is considered of public interest, the municipality conducts online and in-person outreach (targeting communities with limited online access). Community feedback is then incorporated into the Data Usage Protocol.
[13] Data governance	Data transparency	San Jose CA Open Data Portal: the city's open data portal. It aims to maximise available public resources, making the data generated or stored by public bodies accessible, free and usable to all.
	Data protocol	Data Usage Protocol: protocol for medium-high risk systems to govern the collection, access, processing and sharing of data around and ensure compliance with the city's Digital Privacy Policy (Step 5 of the Review process).

Case study 5: Dubai

Population: 3,051,000 (2023)	**Income per capita:** €47,995 (2023)	**Region:** Middle East and North Africa

AI GOVERNANCE FRAMEWORK		
[1] Principles, strategies and guidelines	AI blueprint and roadmap	AI Roadmap (2024) is part of the emirate's Dubai Universal Blueprint for Artificial Intelligence and supports the goals of the Dubai Economic Agenda D33
	Local AI principles	AI Ethics Principles and Guidelines explained below: • AI Ethics Principles: (1) ethics (fair, accountable, transparent and explainable); (2) security (safe and secure); (3) humanity; (4) inclusiveness • AI Ethics Guidelines, make AI systems: (1) fair; (2) accountable; (3) transparent; (4) explainable
[3] Transparency and explainability mechanisms	Municipal website disclosing all AI relevant information	Municipal accessible portal disclosing all AI relevant information and resources, such as landmark projects and strategic documents.
[5] Audits and regulatory inspection	Self-assessment tool	AI ethics self-assessment tool: built to enable AI developer organisations or AI operator organisations to evaluate the ethics level of an AI system, using Dubai's AI Ethics Guidelines.
[11] Promotion of local innovation, knowledge & experimenta-tion	Innovative centres, hubs and laboratories	AI Lab: established in partnership with IBM, it works with a growing network of partners from across governmental and private sectors. Leads Dubai's AI Roadmap.
[13] Data governance	Data transparency	Dubai Pulse: the city's open data portal. It aims to maximise available public resources, making the data generated or stored by public bodies accessible, free and usable to all.
	Data sharing mechanisms	Data sharing toolkit: provides guidance and resources for individuals and private and public organisations to prepare for and design a data-sharing initiative.
	Data privacy	Synthetic Data Framework: designed to aid organisations adopt AI technology, preventing any violation of privacy.

Case study 6: Singapore

Population: 5,918,000 (2023)	**Income per capita:** €79,996 (2023)	**Region:** East Asia and Pacific

AI GOVERNANCE FRAMEWORK		
[1] Principles, strategies and guidelines	AI strategy	National AI Strategy 2.0 (includes an AI playbook).
	Local AI principles	Model AI governance framework's guiding principles: (1) explainability, transparency and fairness; (2) human-centric solutions. AI Verify governance principles: (1) transparency; (2) explainability; (3) repeatability/reproducibility; (4) safety; (5) security; (6) robustness; (7) fairness; (8) data governance; (9) accountability; (10) human agency and oversight; (11) inclusive growth, societal and environmental well-being.
	Guidelines, playbooks and manuals	Public Sector AI Playbook: a resource from the AI strategy for the government. The playbook explains AI, displays common applications in the public sector, provides steps on how to start an AI project and how to develop municipal AI capabilities. Model AI Governance Framework (2nd edition): provides detailed and readily implementable guidance on how to translate ethical principles into practical recommendations that organisations can adopt to deploy AI responsibly. Model AI Governance Framework for Generative AI Advisory Guidelines on Use of Personal Data in AI Recommendation and Decision Systems: done including public consultations. *Other sectorial guidelines:* AI in Healthcare guidelines: provides recommendations to encourage the safe development and implementation of AI medical devices and other AI implemented in healthcare. A Guide to Job Redesign in the Age of AI
[3] Transparency and explainability mechanisms	Municipal website disclosing all AI relevant information	Municipal accessible portal disclosing all AI relevant information and resources, such as landmark projects and strategic documents.
[8] External/ independent oversight and advisory bodies	External advisory and oversight body	Advisory Council on the Ethical Use of AI and Data: comprising 11 multidisciplinary members, it assists the government in developing ethics standards and governance frameworks, and publishes advisory guidelines, practical guides and codes of practice for voluntary adoption by industry.
[9] Alliances, communities of practice and learning groups	Multistakeholder AI coalition	AI Singapore: brings together research institutions and the business ecosystem to research on trustworthy AI and ethical governance, create open-source tools and develop talent for Singapore's AI efforts. Veritas consortium: comprising industry partners and the governmental Monetary Authority of Singapore, it aims to enable financial institutions to evaluate their AI- and data-driven solutions against the principles of fairness, ethics, accountability and transparency.
	Community of practice member	AI Verify Foundation: a global open-source community that convenes AI owners, solution providers, users and policymakers to build trustworthy AI.
[10] Capacity-building programmes	Municipal capacity-building	Municipal staff training: customised for different types of municipal users, Singapore has a directory of courses to achieve various competencies. Included in the Public Sector AI playbook.

[11] Promotion of local innovation, knowledge and experimentation	Innovative centres, hubs and laboratories	Centre for AI and Data Governance (CAIDG): interdisciplinary research centre with multistakeholder partnerships, from governmental agencies to intergovernmental organisations, corporations, academia, think tanks, NGOs and CSOs.
	Regulatory sandboxes	Sandboxes: GenAI Sandbox; Privacy Enhancing Technologies Sandboxes
[13] Data governance	Data transparency	Singapore's open data portal: the city's open data portal. It aims to maximise available public resources, making the data generated or stored by public bodies accessible, free and usable to all.
[14] Other policies and measures	Testing frameworks and toolkits	A.I. Verify: an AI governance testing framework and software toolkit for companies that validates the performance of AI systems against a set of internationally recognised ethical principles through standardised tests. Implementation and Self-Assessment Guide for Organisations (ISAGO): helps organisations assess the alignment of their AI governance practices with the Model Framework. Veritas open-source toolkit: enables the responsible use of AI in the financial industry.
	Green marks	Green Mark for Data Centres Roadmap: charts a sustainable pathway for the continued growth of data centres in Singapore. Green Mark for Data Centres Scheme: for operators that have successfully deployed green data centre best practices, demonstrating superior sustainability and environmental performance.

CONCLUSIONS. POLICY MECHANISMS, CHALLENGES AND RECOMMENDATIONS IN URBAN AI

Marta Galceran-Vercher
Senior Research Fellow, Global Cities Programme, CIDOB

Alexandra Vidal D'oleo
Research Fellow and Project Manager, Global Cities Programme, CIDOB

Local governments worldwide are increasingly adopting algorithmic systems to improve the delivery of public services. However, growing evidence indicates that these systems can cause unintended harms and demonstrate a lack of transparency in their implementation. As a result, the adoption of algorithmic systems has often been accompanied by the development of guiding principles for the responsible use of AI technologies, primarily at the national, supranational or global levels. Notable examples include the OECD AI Principles (2019), the G20 AI principles (2019), the Council of Europe AI Convention drafting group (2022-2024), the Global Partnership on Artificial Intelligence Ministerial Declaration (2022), the G7 Ministers' Statement (2023), the Bletchley Declaration (2023), the Seoul Ministerial Declaration (2024), the EU AI Act (2024) or the UN Report "Governing AI for Humanity" (2024). Yet these frameworks generally provide only broad guidance on what constitutes responsible AI use, offering little practical direction on how these principles should be applied in real-world contexts.

In response to these challenges, many governments are turning to regulatory frameworks and policy tools to operationalise these principles. These efforts are fast emerging, but they vary significantly in scope and approach. Moreover, much of the existing analysis of public sector policy tools tend to focus on national-level perspectives (e.g. OECD, 2024), often overlooking the unique context and challenges faced by local governments.

This CIDOB Monograph has sought to fill this gap by identifying the main policy mechanisms and frameworks leveraged by local governments to ensure that their adoption of algorithmic systems aligns with core ethical principles such as transparency and accountability, fairness and non-discrimination, privacy protection and sustainability. This analysis is complemented by a series of case studies that illustrate how leading cities are implementing these policy mechanisms in practice, resulting in comprehensive local AI frameworks.

This CIDOB Monograph has identified the main policy mechanisms and frameworks leveraged by local governments to ensure that their adoption of algorithmic systems aligns with core ethical principles.

In this concluding chapter, we provide a categorisation of the policy mechanisms identified throughout this publication, along with insights into which mechanisms are most commonly employed by local governments and how they align with the aforementioned ethical principles. We also discuss the challenges faced by local governments and offer recommendations for advancing towards a more responsible use of AI systems in urban environments.

1. Categorisation of policy mechanisms

The criteria used to establish this categorisation were based on the primary function and objectives of the policy mechanisms.

Through a comprehensive review of the policy mechanisms presented across the chapters of this CIDOB Monograph, complemented by a literature review of relevant publications on policy mechanisms for local governments and/or public administrations (including reports from the Ada Lovelace Institute, AI Now Institute and Open Government Partnership, 2021; Ben Dhaou et al., 2024; Jordan et al., 2024; United 4 Smart Sustainable Cities, 2024), we have identified 14 distinct categories of policies currently being implemented by local governments worldwide (see Table 1)[1].

The criteria used to establish this categorisation were based on the primary function and objectives of the policy mechanisms. These include providing normative guidance for the development and use of AI systems, assessing the potential risks of algorithms, ensuring public access to information about algorithmic systems and holding these systems accountable.

Other possible criteria could have focused on whether the mechanisms are oriented towards internal administrative processes (e.g. guidelines for municipal staff or the creation of municipal AI commissions) or external-facing actions, such as the publication of public algorithmic registers or the imposition of bans on controversial AI applications. Additionally, our categorisation does not distinguish between AI-specific mechanisms (e.g. an AI strategy) and indirect mechanisms that contribute to ethical AI governance (e.g. data governance strategies, which, while broader in scope, are critical to AI governance due to the importance of data in AI systems).

1. It is important to note that the limited literature on this topic employs a range of names and terms for the various policy mechanisms, and there is no common vocabulary for their core components.

Table 1. Categories of policy mechanisms for a responsible AI governance in cities

CATEGORY	DESCRIPTION AND PURPOSE	POLICY MECHANISMS
[1] Principles, strategies and guidelines	Policy documents that offer non-binding, normative guidance on ethical principles and values for local administrations, outlining general directions for developing and using AI while managing associated risks. Though the format varies, these documents typically identify high-level policy goals and their relevance to the use of algorithmic systems by public agencies. In some cases, they also provide practical guidance for implementing these principles in the design and deployment of such systems. Ultimately, these guidelines establish normative standards that allow agencies and the public to assess the ethical use of algorithmic systems.	• Ethical AI strategies, action plans, agendas, road maps, charters, handbooks, etc. • Definition of local AI ethical principles: declaration and/or endorsement • Guidelines, playbooks and manuals on how to deploy ethical AI • Internal protocols for AI use • Non-AI specific strategic frameworks that have an impact on AI governance (e.g. digitalisation or data governance frameworks)
[2] Local regulations and laws	Tools aimed at establishing standards, laws and regulations ensuring compliance and addressing societal impacts.	• Local regulation and laws (e.g. regulations to ensure the right to justification, etc.) • Legal compliance mechanisms: to ensure compliance with regional, national or supranational normative frameworks • Regulatory standards (e.g. green AI standards, transparency standards) • Adhering to international regulatory standards • Regulation of controversial AI application: bans, moratoria, etc.
[3] Transparency and explainability mechanisms	Mechanisms for establishing public access to information about algorithmic systems and processes. They are aimed at providing information about algorithmic systems to the general public (e.g. affected individuals, media or civil society) so that they can learn about the use of these systems and demand explanations and justifications related to such use. These mechanisms can function independently or as part of broader frameworks for algorithmic accountability. It is important to distinguish public transparency mechanisms from rights to hearing and explanation, which grant individuals the right to an explanation of specific algorithmic decisions made about them.	• Public algorithm registries • Municipal website disclosing all AI relevant information • Algorithmic transparency standards • Municipal directory of procured AI tools for internal use • Requirements for source code transparency • Explanations of algorithmic logics
[4] Algorithmic impact assessment	Policy instrument used by public agencies to evaluate the potential risks and harms of algorithmic systems. These assessments aim to understand, categorise and address the possible negative effects of algorithms before or during their deployment. Algorithmic impact assessments (AIAs) build on established frameworks from other fields, such as environmental impact assessments, human rights and data protection impact assessments (DPIAs).	• Risk assessment and management procedures (including bias analysis) • Human rights impact assessments • Environment impact assessment mechanisms
[5] Audits and regulatory inspection	Audits encompass a range of practices aimed at examining how a specific algorithmic system functions. Their primary goal is to understand the system's operations and assess its performance against predefined normative standards. While audits share similarities with algorithmic impact assessments (AIAs), they have a distinct time context and are usually conducted either during or after the system's implementation. In contrast, AIAs are typically carried out before or during deployment. Audits can be performed by internal, external or third-party entities, depending on the scope and nature of the assessment. In a third-party audit, an external organisation evaluates the system based solely on its outputs. A second-party audit is conducted by an external assessor who is granted access to both the system's back end and its outputs. First-party audits are carried out by internal members of the organisation.	• Audits of algorithmic systems • Process evaluation • Self-assessment tools
[6] Human oversight, accountability, hearing and appeal procedures	Mechanisms for overseeing and holding AI systems accountable. More precisely, mechanisms that require that decisions made with the assistance of algorithmic systems follow specific procedures designed to safeguard fairness and provide avenues for individuals to seek redress in cases of biased or erroneous outcomes. These procedural safeguards create opportunities for affected individuals or groups to challenge or contest decisions that impact them.	• Internal monitoring and reporting • Human-in-the-loop requirements • Feedback and objection procedures accessible for citizens • Duties of notice of the decision and hearing to the affected parties • Duties to provide reasoned decisions and explanations of a decision • Mechanisms to ensure the right of affected parties to present evidence, appeal and challenge automated decisions

[7] Procurement conditions	Rules governing the acquisition of algorithmic systems by governments and public agencies are crucial for ensuring their accountable use. Many algorithmic systems used by governments are outsourced to private vendors, either through product purchases or service contracts. As a result, vendors play a significant role in the design and deployment of these systems. Terms of procurement contracts are vital in shaping the development and implementation of these systems. Specific procurement conditions, such as requirements for transparency and non-discrimination, can be applied to ensure that the systems acquired meet ethical standards and are used responsibly.	• Procurement clauses • Internal guidelines, frameworks and protocols for AI procurement
[8] Advisory and oversight bodies	Independent oversight bodies, which are intended to oversee and direct the use of algorithmic systems by public agencies. These independent oversight mechanisms are intended to ensure accountability by monitoring the actions of public bodies, and making recommendations, sanctions or decisions about their use of algorithmic systems.	• Advisory bodies: councils, committees, boards, networks, groups, etc.
[9] Alliances, communities of practice and learning groups	Mechanisms aimed at fostering cooperation and partnerships at local, national and international levels.	• Local/national/international learning communities of practice: city networks, working groups, etc. • Local/national/international multistakeholder AI collaborations: networks, platforms, coalitions, etc. • Public-private collaborations and partnership
[10] Capacity-building initiatives	Mechanisms to enhance knowledge and build skills around ethical artificial intelligence. These initiatives can target municipal staff involved – either directly or indirectly – in the design, deployment or use of algorithmic systems, as well as the general public, to promote informed citizenship and encourage broader understanding of AI ethics.	• Municipal staff training (socio-technical approach) • Municipal AI team creation • Municipal AI body: body or cross-cutting committee coordinating/overseeing municipal use of AI • Multidisciplinary approach: creation of diverse teams
[11] Promotion of local innovation, knowledge and experimentation	Mechanisms that provide space for experimentation, innovation and testing in real-world environments.	• Promotion and collaboration with local AI innovation centres, hubs and laboratories • Local AI observatories • Local AI regulatory sandboxes • Initiatives to promote and support local AI ecosystems
[12] Community engagement	Tools to involve citizens, communities and stakeholders in AI decision-making processes; fostering discussions, debates and ensuring that AI policies reflect public concerns and input.	• Public engagement: participatory processes, participatory government models, public listening sessions, promotion of spaces for reflection and debate, communication channels with the public, etc. • Public education (digital literacy) • Local AI ethics boards
[13] Data governance		• Data transparency measures such as open data portals • Data sharing mechanisms • Data rights • Data usage protocols • Data privacy: data privacy laws, synthetic data frameworks, etc. • Data governance systems
[14] Other policies and measures		• Testing frameworks and toolkits • Fiscal incentives such as tax credits, subsidies, etc. • Workforce reskilling programmes • Rating frameworks (e.g. AI star rating frameworks, green marks, etc.)

Source: Authors

2. Alignment of policy mechanisms with ethical principles

The analysis of the policy mechanisms identified throughout this publication[2] reveals that several typologies can be established regarding their alignment with specific ethical principles (see Table 2), which are discussed below.

The analysis of the policy mechanisms identified throughout this publication reveals that several typologies can be established regarding their alignment with specific ethical principles.

a) **Policy mechanisms that serve to uphold particular ethical principles.** For instance, an environmental impact assessment primarily serves sustainability purposes, while a human-in-the-loop measure may uphold both the principle of accountability and fairness by guaranteeing someone oversees the correct functioning of an AI system, ultimately leaving the final decision to a human.

b) **Policy mechanisms that uphold all ethical principles.** Due to their cross-cutting nature or through customisation, some policy mechanisms can advance all ethical principles collectively. For example, policy mechanisms such as AI strategies can be customised to include all ethical principles. Similarly, an oversight committee can be tasked with overseeing privacy protection, accountability or the full range of ethical principles.

Despite their potential, not all are frequently applied by cities striving to establish ethical local AI frameworks. Some noteworthy examples among these mechanisms are:

- **Principles, strategies and guidelines**: one of the most frequently applied mechanisms by cities worldwide. Cities are consistently implementing these policy mechanisms from an ethical standpoint to provide them with a base and sense of direction. These mechanisms are particularly used by municipalities to demonstrate their willingness to commit to responsible AI deployment.

- **Procurement clauses**: since municipalities often lack the resources to develop their own in-house AI systems, another commonly applied policy mechanism are procurement clauses. They become essential and a practical go-to solution. They enable municipalities to leverage their purchasing power when acquiring AI systems while promoting ethical AI development by private sector providers.

- **Outward-facing mechanisms**: more and more cities are relying on **advisory and oversight bodies** consisting of external and independent experts who advise the municipality on ethical conundrums and oversee their use of algorithmic systems. Similarly, many municipalities are engaging in **alliances, communities of practices and learning groups** to jointly address challenges and identify ways in which to use AI safely.

- **Data governance**: while data governance may not be immediately perceived as a direct policy mechanism for AI, it lays the foundation for a correct management and safeguarding of citizens' data, and is crucial for non-discriminatory systems, making it a vital component of an ethical deployment of AI. Data governance serves as a building block, enabling data transparency, data rights protection, data privacy and the sustainable use of data

2. Extracted solely from the chapters (Part I) and case studies (Part II) of this monograph.

for AI systems. Examples include protocols for the anonymisation of personal data, or the use of synthetic data to train AI systems, in order to solve privacy and fairness concerns.

- **Audits**: in spite of being universally recognised by experts and civil servants as one of the most critical policy mechanisms for safeguarding ethics, audits remain underutilised. Their infrequent use is largely due to constraints imposed by private AI providers, and a lack of in-house technical capacity.

c) **Policy mechanisms not specifically associated to any ethical principle**. While not tied to any specific ethical principles, these policy mechanisms are considered key for an ethical AI deployment, nonetheless. They establish structured processes to be followed; coordinate its deployment; or provide the necessary expertise and knowledge for informed decision-making. Some noteworthy examples of these mechanisms are:

- **Internal protocols for AI use**: most cities develop internal protocols to guide their use of AI, providing step-by-step structures that facilitate its implementation by the municipality. Some cities, albeit a few, additionally complement them with comprehensive **protocols for AI procurement**. These protocols can then include mandatory impact assessments, bias analysis and other policy mechanisms to ensure respect for specific ethical principles.

- **Innovative AI centres, hubs and laboratories:** a significant number of cities collaborate, promote or have established innovative AI centres, hubs and laboratories to create practical knowledge and develop AI solutions. The research conducted by these institutions is oriented from an ethical standpoint.

- **Capacity-building initiatives:** one of the least commonly implemented mechanisms is the creation of dedicated **municipal AI teams** with the expertise to audit or develop in-house AI systems. This is primarily due to technical and financial constraints on the part of municipalities. In contrast, many have established **municipal AI bodies** tasked with coordinating AI use across departments. These bodies play a critical role in facilitating the cross-cutting monitoring of AI deployment within the municipality, ensuring a more organised and accountable approach to AI governance.

Table 2. Policy mechanisms alignment with ethical principles

POLICY MECHANISMS (PM)		ETHICAL PRINCIPLES			
		Accountability and transparency	Privacy and data governance	Fairness and non-discrimination	Sustainability
[1]	AI strategies	x	x	x	x
	Local AI ethical principles	x	x	x	x
	Guidelines, playbooks and manuals	x	x	x	x
	Internal protocol for AI use				
[2]	Local regulations and laws	x	x	x	x
	Legal compliance mechanisms	x	x	x	x
	Regulatory standards	x	x	x	x
	International regulatory standards	x	x	x	x
	Regulation of controversial AI application		x	x	
[3]	Public algorithm register	x	x	x	
	Municipal website disclosing all AI relevant information	x			
	Municipal directory of procured AI tools for internal use	x			
[4]	Risk assessment and management	x	x	x	x
	Human rights impact assessments	x	x	x	x
	Environmental impact assessment				x
[5]	Audits	x	x	x	x
	Self-assessment tools	x	x	x	x
[6]	Internal monitoring and reporting				
	Human-in-the-loop	x		x	
	Feedback and objection procedures accessible for citizens	x		x	
[7]	Procurement clauses	x	x	x	x
	Internal protocols for AI procurement				
[8]	Advisory and oversight bodies	x	x	x	x
[9]	Alliances, communities of practice and learning groups	x	x	x	x
[10]	Municipal staff training	x	x	x	x
	Municipal AI team				
	Municipal AI body				
	Multidisciplinary approach			x	
[11]	Innovative AI centres, hubs and laboratories				
	Local AI observatories				
	Regulatory sandboxes		x		
[12]	Public engagement	x		x	x
	Public education (digital literacy)		x	x	x
	Local AI ethics boards	x	x	x	x
[13]	Data governance	x	x	x	x
[14]	Testing frameworks and toolkits	x	x	x	x
	Fiscal incentives (tax credits, subsidies, etc)				x
	Workforce reskilling programmes				x
	Rating frameworks	x	x	x	x

Source: Authors
Table legend: Yellow (PM aligned with a specific ethical principle or several simultaneously); Blue (cross-cutting or customisable PM that serves all ethical principles); Green (PM not associated to a specific ethical principle but relevant for a responsible deployment of algorithmic systems in general). Dark grey (PM explicitly mentioned in the chapters of Part I, see Annex I); Light grey (PM not mentioned in the chapters of Part I).

Regardless of their level of specificity, scope or effectiveness, policy mechanisms are not applied evenly by cities, varying in frequency due to resource constraints, technical limitations or differing local priorities. Table 3 offers a comparison of the cities included in the case studies, highlighting how frequently certain policy mechanisms are applied, which ones are most commonly employed, and which ones are rarely called upon. While we thought it relevant to offer a comparison of the cities analysed in Part II of this publication, it should be acknowledged that the conclusions drawn from the selected case studies are limited by the small size of the sample.

Table 3. Case studies comparison (most used policy mechanisms)

POLICY MECHANISM		CITIES					
		Barcelona	Amsterdam	New York	San José	Dubai	Singapore
[1]	AI strategies, agendas, action plans, handbooks, road maps, etc.	x	x	x	x	x	x
[1]	Local AI ethical principles	x	x	x	x	x	x
[3]	Municipal website disclosing all AI relevant information	x	x	x	x	x	x
[13]	Open data portal (data transparency)	x	x	x	x	x	x
[1]	Guidelines, playbooks and manuals		x	x	x	x	x
[8]	Advisory and oversight bodies	x	x	x	x		x
[9]	Community of practice member	x	x	x	x		x
[1]	Internal protocols for AI use	x	x	x	x		
[4]	Risk assessment and management	x	x	x	x		
[7]	Procurement clauses	x	x	x	x		
[12]	Public engagement	x	x	x	x		
[3]	Public algorithm register	x	x		x		
[4]	Human rights impact assessment	x	x		x		
[9]	Multistakeholder AI collaborations		x		x		x
[10]	Municipal staff training (municipal capacity-building)	x	x				x
[10]	Municipal AI body	x		x	x		
[11]	Innovative AI centres, hubs and laboratories		x			x	x
[13]	Other data governance policies (data rights, data sharing mechanisms, data protocols, etc.)	x			x	x	
[2]	Algorithm transparency standard	x	x				
[6]	Internal monitoring and reporting			x	x		
[6]	Feedback and objection procedures accessible for citizens		x		x		
[5]	Audits	x	x				
[2]	Regulation of controversial AI application			x			
[3]	Municipal directory of procured AI tools for internal use			x			
[10]	Municipal AI team creation (municipal capacity-building)		x				
[11]	Local AI observatory	x					
11]	Regulatory sandboxes						x
[14]	Testing frameworks and toolkits						x
[14]	Rating frameworks						x

Source: Authors.
Note: The list of policy mechanisms has been ordered first, by most to least frequent; second, by categorisation.

3. Challenges and recommendations

a. There are few references on how to effectively operationalise ethical AI principles at the local level. Most existing guidelines, studies and capacity-building programmes fail to account for the unique challenges faced by urban administrations, which are often disconnected from national strategies and policies. This gap is further compounded by the heterogeneous nature of local governments, which vary widely in terms of size, resources and capabilities.

Regardless of their level of specificity, scope or effectiveness, policy mechanisms are not applied evenly by cities, varying in frequency due to resource constraints, technical limitations or differing local priorities.

Recommendations:

- Localise (i.e. attune to local context) regional, national or global ethical principles and AI governance policy mechanisms by creating local definitions of success and identifying local priorities.

b. Local administrations face a significant shortage of talent and technical expertise, a challenge that is further compounded by the global scarcity of AI professionals, making it difficult to attract qualified individuals at the local level. As a result, municipal governments often have limited understanding of the potential impacts and implications of algorithmic systems.

Recommendations:

- Prioritise capacity-building programmes as part of municipal strategies and policy frameworks for governing and regulating algorithmic systems. This should include allocating specific resources for municipal training programmes, investing in public awareness campaigns and promoting initiatives to build foundational knowledge and skills around ethical AI within public administration.

- To overcome the challenges of attracting local talent, local governments should invest in strategies that facilitate the exchange and adaptation of knowledge from local stakeholders. Additionally, establishing strong alliances and connections with knowledge-sharing networks can help bridge expertise gaps.

- Adopt a holistic approach to capacity-building by encouraging public debates and awareness-raising initiatives within local communities. These efforts should focus on educating citizens about the opportunities and risks associated with the use of algorithmic systems.

c. Ensuring the **transparency and accountability of algorithmic systems** used by local governments presents several challenges, including managing public perception and potential backlash regarding external-facing AI systems, adapting organisational culture and work practices for internal-facing AI systems, and fostering shared ownership across the public administration (i.e. AI should not be seen as the sole responsibility of the IT department).

Recommendations:

- Embed transparency as a core objective beyond just AI-specific policies. This includes fostering a culture of transparency and accountability throughout the entire AI life cycle.
- Encourage the use of open-source code, which can enhance trust and allow for greater scrutiny and collaboration.
- Clarify responsibilities by designating a specific point of contact or "AI project lead" for all AI initiatives, ensuring accountability and streamlined communication.
- Create multiple feedback channels and integrate evaluations at key stages of the project to ensure continuous improvement and responsiveness.
- Allocate budget for comprehensive explanatory phases, ensuring that stakeholders, both internal and external, fully understand the AI systems and their implications.

d. Ensuring **privacy protection** presents specific challenges, including a complex regulatory landscape that local governments often struggle to navigate. Data-related issues are closely tied to policy mechanisms designed to safeguard data protection. Notably, there is a limited availability of high-quality data in urban environments, which can be attributed to several factors: inadequate data management practices, ethical concerns and risks surrounding the large-scale collection of data and poor data sharing between administrations due to the absence of unified standards and underdeveloped data governance frameworks.

Recommendations:

- Generate high-value public datasets by improving data collection and management practices to enhance data quality and utility.
- Promote interoperability and collaboration across agencies and sectors to facilitate seamless data exchange and sharing.
- Create secure and transparent frameworks for data sharing that ensure privacy protection while enabling innovation.
- Encourage innovation and experimentation within controlled environments, such as regulatory sandboxes, to test new data-driven solutions safely and responsibly.

e. Algorithmic systems may reinforce existing urban inequalities while creating new forms of discrimination, hence the importance of considering the **notion of fairness and non-discrimination** when local administration deploy AI systems. A specific challenge in this regard includes the fact that discrimination automated by AI is more abstract, opaque, difficult to detect (black boxes) and large-scale. Hence, it disrupts traditional legal remedies and procedures usually employed by local governments for detecting, preventing and correcting it.

Recommendations:

- Consider the multiple roles of public administrations – as developers, deployers and regulators – when designing initiatives to enhance the fairness and non-discrimination of algorithmic systems.
- Localise existing policy frameworks to address unintended discrimination in algorithmic systems, ensuring they are tailored to the unique challenges of urban environments.
- Embed a holistic approach to AI governance within local administrations, considering the socioeconomic impacts throughout the entire AI life cycle, from design to deployment.
- Ensure diversity among the teams involved in the design and deployment of algorithmic systems to reduce the risk of biased outcomes and promote inclusive solutions.

f. The main challenge associated with the **environmental sustainability of AI** principle is related to the fact that AI for sustainability often clashes with the sustainability of AI. At the same time, there are few frameworks for the sustainability of AI with an urban focus.

Recommendations:

- Assess the full life cycle impact of AI systems to determine whether their benefits outweigh their environmental costs. Minimise the carbon footprint of city-wide AI deployments by prioritising energy-efficient systems, adopting green computing practices, utilising Tiny ML and powering data centres with renewable energy.
- Foster a circular economy around data centres by reducing electronic waste. Promote responsible sourcing, reusing and recycling of AI hardware. Encourage the reuse and recycling of AI technologies, data and infrastructure.
- Repurpose the energy and resources used by AI infrastructure and deploy AI systems that integrate seamlessly with existing urban infrastructure, optimising both energy use and system efficiency.

Finally, it is important to acknowledge the constraints of the research presented in this CIDOB Monograph. The study was limited by both its relatively short time frame and the challenges inherent in collecting information within the context of GOUAI (see Galceran-Vercher and Vidal, 2024). As a result, we recognise that some key examples of algorithmic policy mechanisms and governance frameworks may not be included. Furthermore, most of our evidence is drawn from policies promoted by the Global North, primarily through interventions led by local governments in the United States and Europe. This geographic focus is another limitation that future research within the GOUAI context will aim to address. We acknowledge that a more systematic analysis of governance policies and practices from the Global South could provide new insights, revealing different policy approaches, priorities and implementation challenges.

References

Ada Lovelace Institute, AI Now Institute and Open Government Partnership. "Algorithmic Accountability for the Public Sector", 2021

Ben Dhaou, S., *et al*. "Global Assessment of Responsible Artificial Intelligence in Cities: Research and recommendations to leverage AI for people- centred smart cities". Nairobi: United Nations Human Settlements Programme (UN-Habitat), 2024

Jordan, C.; Glickman, J. and Panettieri, A. "AI in cities: Report and Toolkit". Washington: National League of Cities and Google, 2024

OECD. "OECD AI Policy Observatory: Catalogue of Tools & Metrics for Trustworthy AI", 2024

United 4 Smart Sustainable Cities. "Guiding principles for artificial intelligence in cities". Geneva: International Telecommunications Union, 2024

ANNEX 1. List of policy mechanisms mentioned in the ethical principles chapters (Part I)

Ethical Principle	Policy mechanisms
Accountability and transparency	• Impact assessments [4] • Procurement clauses [7] • External algorithmic audits [5] • Algorithm registers [3] • Transparency standards [3] • Interdisciplinary governance oversight committees [8] • Participatory processes (throughout the AI life cycle) [12] • Human-in-the-loop design [6] • Civil servants' education [10] • Connect with knowledge-sharing networks [9] • Local stakeholder collaboration [9]
Privacy and data governance	• Legal compliance [2] • Risk management systems [4] • Data governance systems [13] • Impact assessments [4] • Auditing [5] • Algorithm repositories and AI registers [3] • Regulatory sandboxes [11]a • Urban AI strategies [1] • Multistakeholder collaboration [9]
Fairness and non-discrimination	• AI strategies [1] • Risk analysis and protective mechanisms [4] • Impact assessments [4] • Local AI standards for fair AI [2] • Procurement standards for fair AI [7] • Urban laws for the right to justification [2] • Multidisciplinary advisory bodies [8] • Diverse and interdisciplinary teams • Audits [5] • Mitigation techniques in AI life cycle [14] • Knowledge sharing networks [9] • Municipal training [10] • Public education [12] • Ethical principles [1] • Bias analysis [4] • Digital rights protection [14]
Sustainability	• Environmental, social and governance standards (e.g. green AI standards) [2] • Impact assessments [4] • Monitoring and auditing [5] • AI for sustainable cities consortiums [9] • Fiscal incentives: tax credits or subsidies [14] • Local AI ethics boards [12] • Urban data privacy laws [13] • Public engagement [12] • Participatory government models [12] • Digital literacy campaigns [12] • Workforce reskilling programmes [14] • AI energy star rating frameworks [14] • International collaboration [9] • International sustainability standards [2]

Source: Authors
Note: The policy mechanisms described here preserve the original wording from the ethical principles articles (Part I) and have been categorised according to the authors' categorisation provided in the concluding chapter of the Monograph. The mechanisms may not be listed under the "policy mechanism section" of the articles but may be found throughout the article itself.

Artificial intelligence (AI) promises to revolutionise urban spaces, offering solutions to the most significant urban challenges humanity faces today. Cities serve as ideal testing grounds for AI, with local governments adopting data-driven technologies to automate routine tasks, improve efficiency and make cost-effective decisions. However, the rapid deployment of algorithmic systems raises ethical concerns, particularly regarding citizens' rights and environmental and social impacts. Responsible AI governance is critical to avoid unintended negative consequences while reaping the benefits this technology can deliver. Cities that are already implementing ethical AI policies can provide valuable examples for others, highlighting the need for local policymakers to adopt approaches that prioritise ethical, transparent and inclusive AI deployment.

With that in mind, CIDOB presents this monograph, exploring specific policy mechanisms and existing governance frameworks that promote responsible urban AI on the ground. First, it analyses how essential ethical principles – namely (1) accountability and transparency; (2) privacy and data protection; (3) fairness and non-discrimination; and (4) sustainability – can be effectively implemented in urban environments through targeted policy measures. Then, it presents case studies of cities worldwide that have established comprehensive AI governance frameworks. Ultimately, this CIDOB Monograph aims to serve as a practical document that can inspire action and guide other local public sector actors on the long road ahead to guaranteeing ethical deployment of urban AI.