PART I. OPERATIONALISATION OF ETHICAL PRINCIPLES IN URBAN SETTINGS

ACCOUNTABILITY AND TRANSPARENCY IN URBAN AI

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. In these ways, transparency becomes an enabling condition for developing algorithmic accountability by providing ways forward for contestation.

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.

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.

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).

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).

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.

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.

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.

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.

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