CIDOB briefings

ISSN 2564-8713

33
NOVEMBER
2021

AI ETHICS IN POLICY AND ACTION: city governance of algorithmic decision systems

Andrea G. Rodríguez, Researcher & Project Manager, Global Cities Programme, CIDOB (Barcelona Centre for International Affairs).

In July 2021, CIDOB, alongside the cities of Barcelona, Amsterdam and London, set up the Global Observatory of Urban Artificial Intelligence (GOUAI). The Observatory's first output is the publication of a framework for the analysis of city initiatives that involve the use or development of artificial intelligence systems (AI). On July 15th 2021 CIDOB hosted a working session with representatives from the Cities Coalition for Digital Rights, UN-Habitat and Barcelona City Council. This CIDOB Briefing introduces the framework and builds on the conclusions of the session.

1. Introduction

As digitalisation advances, cities are increasingly embracing the use of algorithmic tools to improve efficiency when allocating resources, face structural challenges such as climate change and deliver better and faster services to citizens. Some of these benefits are already being felt. The city of Fuengirola in Spain uses artificial intelligence (AI) to improve public health by monitoring the capacity of its beaches, allowing local authorities to prevent overcrowding. The cities of New Jersey (USA) and Stara Gora (Bulgaria), meanwhile, are testing smart traffic light systems to prevent congestion, improve road safety and reduce pollution.

In spite of the benefits of using artificial intelligence, the wide array of future applications for cities has sparked debate about their ethical implications. Algorithmic tools can reproduce the society's biases and result in indirect discrimination, especially of vulnerable groups. Algorithms used by law enforcement authorities have been proven to have lower accuracy rates when applied to racial minorities. COMPAS, used in US court systems to prevent recidivism or reengagement in criminal behaviour, is one example. Others, such as the algorithm used to assist Amazon's hiring process, have been found to have a gender bias and discriminate against women.

For these reasons, the civil organisations AlgorithmWatch and Access Now issued a joint declaration calling for the establishment of rigorous transparency mechanisms and the creation of public registries of algorithms used by public authorities. Cities have entered the debate too.

In the 2018 Declaration of Sharing Cities a number of cities agreed to implement ethical standards in their use of digital technologies and to protect citizens' rights to privacy, security and digital identity. Later, the Cities Coalition for Digital Rights (CC4DR) developed these principles and agreed to foster a democratic and inclusive process of digitalisation by calling for universal access to the internet, increased citizen participation and the enforcement of ethical principles such as the right to privacy and algorithmic transparency and non-discrimination.

It was in this light that in July 2021 the CC4DR, CIDOB and the cities of Barcelona, Amsterdam and London established the Global Observatory of Urban Artificial Intelligence (GOUAI). At the first working session of the GOUAI, which took place online on July 15th 2021, the coalition cities discussed the ethical principles within which the Observatory would frame its research. The next section of this briefing examines the debates about AI ethics in cities and explores the limitations of these approaches. The third section refers to the methodology used to identify the key guiding principles discussed in the session. Lastly, the briefing ends with a reflection on how cities can move forward in designing and implementing people-centric ethical artificial intelligence.

2. AI ethics for the city

Artificial intelligence provides cities with an opportunity to tackle the growing challenges they face. It can help local authorities model solutions and identify key variables to inform better policies in areas such as climate change and demographic growth, and to identify new social needs and trends early on. This should help them act preventively and mitigate adverse effects, such as predicting peaks of pollution in certain urban areas, to give a public health example.

But as well as benefits, algorithmic tools also pose many risks that cities must address if responsible use is to be ensured. One of these is the perception that the negative effects of using AI are merely technical challenges, neglecting their socio-political implications. In *The Smart Enough City* (2019), Ben Green argues that cities often overestimate the value of technology and do not sufficiently consider the risks. For example, the procurement of AI systems can create power asymmetries between the government and the companies that build the systems that can expose citizens' personal information to undesired parties. This trend worsens as urban networks grow, as Jatham Sadowski points out in *The Spectrum of Control: A Social Theory of the Smart City* (2015). As the attack surface expands, cybersecurity risks grow: the higher the number of interconnected devices, the more points of entry for cybercriminals.

Cities can benefit from their proximity to the citizenry to test solutions, examine implications and establish channels of communication to identify misuse and bad performance.

Both the above examples demonstrate how artificial intelligence can undermine cities' efforts to improve cohesion and reduce inequalities. A 2020 report by the UK government's Centre for Data Ethics and Innovation found that local authorities are increasingly using data science to support their decisions in sensitive sectors such as welfare and social care, healthcare and housing. The input or reproduction of societal biases by algorithmic tools can end up discriminating against vulnerable groups such as women, LGBT people, ethnic minorities and those on low incomes. An example of unfair correlations emerged in Norrtälje (Sweden) in 2019, where the city developed an AI system to identify children at risk that it subsequently decided not to implement, as it risked reproducing social prejudices.

Implementing ethical AI means engaging in a constant debate with the public in order to identify worries and priorities, as well as to establish points of connection in order to navigate trade-offs. In this sense, as forefront implementers of algorithmic solutions that affect the everyday lives of citizens, cities can benefit from their proximity to the citizenry in order to test solutions, examine implications and establish channels of communication to identify misuse and bad performance. But to have a transformative effect and influence how AI is designed, tested and implemented cities need to agree on which values and ethical standards are necessary for the responsible design and implementation of algorithmic tools. These discussions go beyond technical approaches that seek to "encode" new rules of performance or solve deviations.

To respond to these challenges cities like New York have created task forces to address algorithmic discrimination. Others, like Amsterdam and Helsinki, have opened algorithmic registries to increase transparency about the data they use and where they are deployed and to establish communication channels with citizens in order to acquire feedback, in line with the aforementioned recommendation by Access Now and AlgorithmWatch. Urban artificial intelligence strategies are emerging in cities like Barcelona that seek to direct the efforts of local administrations and set limits on use or development, while organisations such as the Cities Coalition for Digital Rights have become spaces for discussing the impact of algorithmic tools on people's digital rights.

3. Minimal ethical standards (MES)

There is no single approach to AI ethics. Different industries and authorities have different priorities (see: Figure 1). Cities themselves have divergent approaches to AI ethics. For example, the AI Strategy published by New York City in October 2021 identified four principles as necessary to implement ethical AI: accountability, fairness, privacy and security, and community engagement and participation. By contrast, the strategy published by Barcelona City Council in April 2021 recognises seven principles: as well as the four defined by New York, it includes environmental sustainability and puts extra emphasis on transparency.

Reaching consensus on the issue of whether or not AI should be regulated and if so which principles should be considered ethical issues is no easy task. For that reason the GOUAI drew up a report for the purpose of creating a **structured framework of minimal ethical standards (MES) that can help city administrators mitigate adverse effects and plausible underperformance due to issues with design, implementation strategy or oversight**.

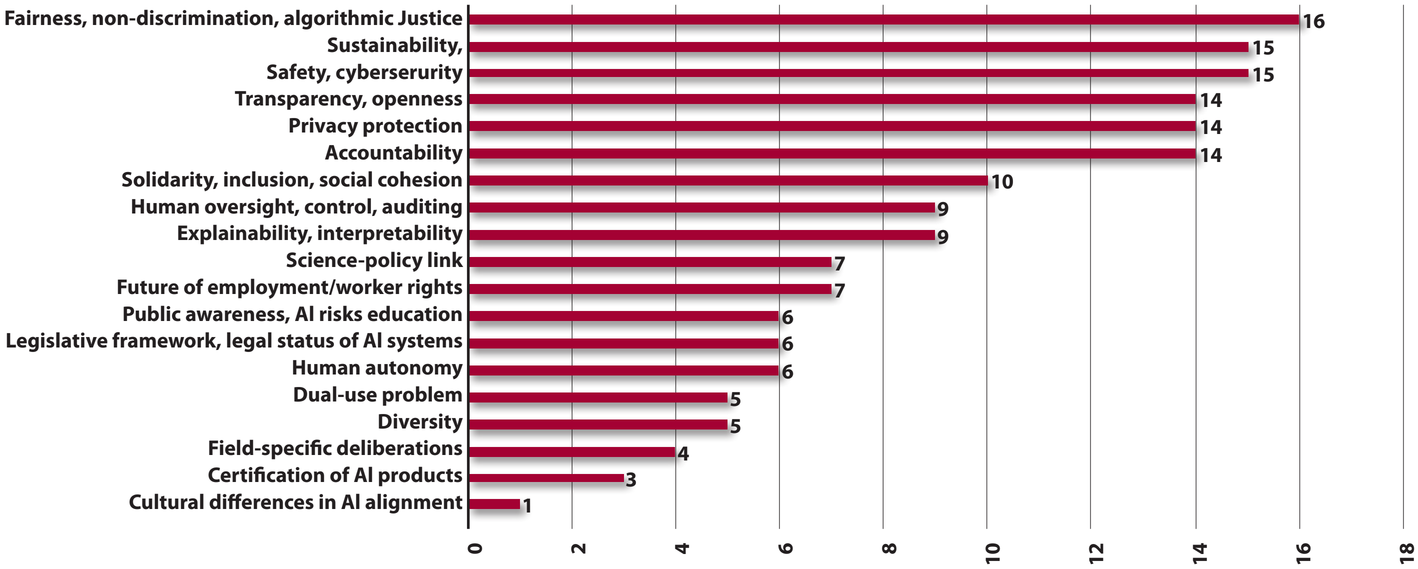
These are *minimal* standards because they aim to guarantee in as few principles as possible that the deployment of AI systems works for the common good and promotes citizens' well-being while at the same time respecting fundamental rights and freedoms. The principles have two objectives: on the one hand they offer a sociotechnical approach to AI ethics that reconciles technical fixes with the social discussion; on the other, the framework advances topics that are necessary for sustainable technological progress, such as environmental concerns.

To identify the MES, the author reviewed 19 key documents that deal with AI ethics from academia, NGOs and the public and private sectors. These documents identified 19 principles as "ethical" in the use or development of artificial intelligence. Six had an average mention rate of 77.19%, followed by three with a mention rate of 49.12% and ten with a mention rate of 26.32%. The first six were the object of the discussion, which was later enriched with the experiences of the various cities and lessons from the other principles.

Fairness and non-discrimination

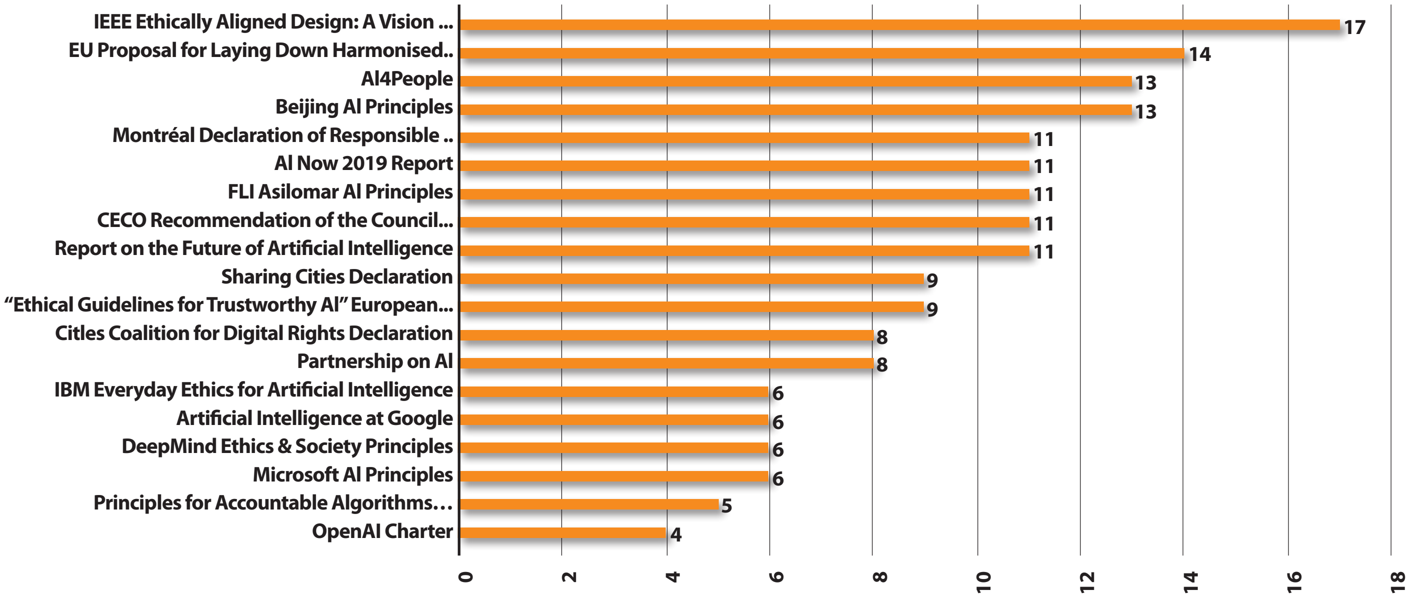
The AI community largely agrees that ensuring fairness is essential, but there is no clear definition of what that means. In fact, ensuring fairness means recognising

Figure 1: Frequency of appearance of identified ethical principles (out of 19)



Source: Prepared by the author.

Figure 2: Number of ethical principles identified per document



Source: Prepared by the author.

which attributes society considers necessary for an algorithm to be *fair*, a debate that echoes those on egalitarianism and social justice in the field of philosophy (Binns, 2018).

Generally, fairness tends to refer to the ability of an algorithmic tool to not discriminate against an individual or group because of personal characteristics or membership of a vulnerable group, such as ethnic or sexual minorities. This approach is largely followed by the AI community, and may be seen in the Declaration of the Cities Coalition for Digital Rights, the OECD's recommendations and the European Union's work on AI ethics.[1]

AI can reproduce existing biases in society inadvertently represented in the training dataset or it can learn them from new inputs. In summer 2021 the city of Nissewaard (Netherlands) stopped using algorithmic tools to automatically allocate social assistance packages because the system was found to be "unreliable" by a private investigation.

An structured framework of minimal ethical standards (MES) that can help city administrators mitigate adverse effects and plausible underperformance due to issues with design, implementation strategy or oversight.

Transparency and openness

Transparency and openness are two interrelated terms that refer to both the technical aspects of algorithmic tools and the socio-political context in which they operate. In that sense it is possible to distinguish between technical transparency and transparency *in use*.

Technical transparency refers to the ability of an external actor to investigate the relationship between two variables and see *through* the architecture of the model, examining the relation between cause and effect (interpretability) and / or the significance of the weighting (explainability). Interpretable and explainable models of AI are as important as determining responsibility.

Socio-political transparency considers that **technical transparency is not enough to guarantee fundamental rights and freedoms.** To do so, it is necessary to guarantee an environment of administrative openness and to improve citizen participation. This notion of "transparency in use" recognises the need to create an environment of trust between the technology, the city and its residents to ensure that the algorithmic tool is used following the principle of proportionality. To guarantee transparency in use, in 2020 the cities of Helsinki and Amsterdam announced the creation of a public algorithm registry that depicts the algorithmic tools used in the city and describes the data used for that purpose. It is also open to public participation, enabling citizens to report bad performance or other concerns.

Safety and cybersecurity

To remain aligned with human values, to "stay" ethical, **algorithmic decision systems must be resilient to cyberattacks and adversarial threats that have the objective of altering their functions,** for example, by adding noise to the training dataset or creating perturbations in new inputs from which the algorithm learns (Akhtar & Mian, 2018). The process of ensuring a high degree of cybersecurity and technical robustness extends beyond the system design phase. In fact, the hardest task is to ensure that the system remains invulnerable after deployment.

In addition to cybersecurity, urban AI must be safe to use. Malperformance can damage not only people's rights but also endanger physical safety. The City of Florence in Italy is developing a smart system to improve safety at tram stops. The system will distinguish between humans and vehicles and will be able to act accordingly to avoid accidents (Iolov, 2021). Errors in design or external threats can make an algorithmic tool an extra concern for cities.

Privacy protection

The protection of privacy and the right to intimacy in the public space are key challenges of the digital era. Privacy has been recognised as one of the most prominent digital rights and an essential component of the ethical use of algorithmic tools. In fact, the Cities Coalition for Digital Rights makes privacy a condition of ensuring human dignity (CC4DR, 2018, principle 2) and asks member cities to put local policies in place and develop tools for citizens to protect their privacy.

Local administrations should pay attention to the level of intrusiveness and assess the potential impact, direct or indirect, of the tool on citizens' right to privacy. Moreover, cities must guarantee the anonymisation of

[1] The various deliverables of the Commission's appointed High-Level Expert Group on Artificial Intelligence define fairness as commitment to equality and non-discrimination. This definition of fairness is the one used in the 2020 AI white paper and in the 2021 Proposal for an Artificial Intelligence Act. Similarly, the Joint Research Centre often refers to "fairness and non-discrimination" as a single concept too.

data and create security against deanonymisation. One example is the adoption of robust encryption algorithms and measures that guarantee individual sovereignty over personal data.

Cities should also guarantee that data is always available and adhere to the right to explanation. An example of good practice can be found in Grand Lyon, where the metropolitan authority promotes a free data wallet service for citizens to securely store official documents, passwords and house bills. A similar example could be brought in for the data used in urban AI applications. The service puts the citizen in control of their personal data, and they can decide with whom to share it.

Cities can additionally benefit from putting in place measures to guarantee data and technological sovereignty, a topic that has gained momentum over the last years (see: Declaration of Sharing Cities) and one of the principles required for membership of the Coalition (CC4DR, n.d.).

Sustainability

Global challenges such as climate change are forcing cities to advance their green agenda. Digitalization can be both an instrument and an obstacle. The environmental dimension of AI is often overlooked when discussing AI ethics, but it is nevertheless central to guaranteeing social wellbeing. For that reason, the framework considers sustainability to be one of the principles that ethical AI must protect.

Algorithmic tools depend on data. The data lifecycle has a severe polluting effect on the environment. Data is stored and often processed in data centres, most of which rely on non-renewable energies to power the computers and cool the facilities. What is more, to gather, use and visualise data cities can end up generating a great deal of electronic waste in the form of sensors, computers and devices.

In Belo Horizonte (Brazil) in 2008 the city established a reconditioning centre to reduce electronic waste. The centre connects disadvantaged communities that cannot afford electronic devices and at the same time helps build necessary digital skills at more than 300 sites that offer free internet and access to computers.

Reliance on data can also foster competitive behaviour between different companies and cities to mine more data to be used to feed new algorithmic tools or improve existing ones. Despite these negative effects, AI can help cities fight climate change by providing new insights and correlations like creating models that help cities more precisely investigate causes of pollution and aiding administrations in better allocating resources.

Cities can benefit from greater control of the data lifecycle, for example, by partnering with the private sector to open urban data centres that commit to the city's sustainability efforts, exercising mindful control over the traceability of the hardware employed and releasing AI policies that contain a green lens. The City of London, for example, follows a responsible procurement policy by which it assesses a contract's viability according to the environmental impact of the product or service (along with the maximisation of social value and respect for the ethical treatment of people) in line with its commitments towards the United Nations' Sustainable Development Goals.

4. Conclusion

The many documents reviewed from the academic, private and public sectors show the growing need to find consensus in the direction AI should take. Given the lack of international regulation and the lack of effective consensus on what is an ethical issue, cities, through organised action, have the chance to find agreement on these topics and endorse a framework of minimal ethical standards that, as they are put into practice, influence the global governance of artificial intelligence.

City action is a necessary step towards a socially responsible governance of emerging technologies that challenges the procedures of local administrations and involves the use of citizen data. The principles described in the previous section of this briefing described the principles that are necessary to guarantee an ethical use of AI-enabled technologies but also include some challenges that cities are already facing.

Significant consensus existed between participants on the importance of the principles of the framework, but also on involving citizens in the lifecycle of the algorithmic tools implemented in cities. Municipal authorities must ensure citizen participation has real impact on the policy cycle around the use of AI and the design of AI policies. Since not all risks are foreseeable, monitoring AI systems after implementation provides an opportunity to engage citizens in the process. Citizen participation is essential both for reporting bad performance and to ensure city officials are informed about public concerns. Amsterdam's public algorithm registry offers a good example (City of Amsterdam, 2020).

As AI applications take over the public space, city officials will find themselves in the position of balancing the enforcement of principles that may be in tension with each other. For example, friction may exist between the principles of privacy protection and fairness when improving levels of representativeness in the training data set. Local governments and private companies may need to access data that is under protection and when doing so efforts must be made to mitigate the risks of algorithmic discrimination.

The cities also agreed that the recognition of a set of minimal ethical standards is not enough to mitigate all the risks of the use of artificial intelligence. Different areas of application require different measures to guarantee MES. The upcoming EU AI Act (COM/2021/206 final) identifies high-risk applications, such as real-time biometric recognition and sectorial applications such as those in health or transportation, and creates extra obligations for providers and implementers. A similar risk-based approach could help cities better navigate trade-offs, mitigate risks and develop AI strategies that respond to the real challenges facing urban areas.