AI and Cities

Risks, Applications and Governance

Mila

UN-HABITAT
FOR A BETTER URBAN FUTURE

Acknowledgements

Lead Authors

Shin Koseki (Université de Montréal)
Shazade Jameson (Tilburg University, Mila)
Golnoosh Farnadi (HEC, Mila)
David Rolnick (McGill, Mila)
Catherine Régis (Université de Montréal, Mila)
Jean-Louis Denis (Université de Montréal)

Authors

Amanda Leal, Cecile de Bezenac, Giulia Occhini, Hugo Lefebvre, Jose Gallego-Posada, Khaoula Chehbouni, Maryam Molamohammadi, Raesetje Sefala, Rebecca Salganik, Safiya Yahaya, Shoghig Téhinian

UN Habitat Reference Group

Cerin Kizhakkethottam, Christelle Lahoud, Herman Jean Pienaar, Ivan Thung, Judith Owigar, Katja Schaefer, Kerstin Sommer, Leandry Nkuidje, Melissa Permezel, Paula Pennanen-Rebeiro-Hargrave, Pontus Westerberg, Remy Sietchiping, Salma Yousry

Mila Core Team

Benjamin Prud'homme (co-lead)
Rose Landry (co-lead), Anna Jahn (reviewer)

UN Habitat Core Team

Abdinassir Shale Sagar (lead)
Melissa Permezel (reviewer)

Editing

Andrea Zanin, Daly-Dallaire, Services de traduction

Design, Layout and Illustrations

Simon L'Archevêque

Contents

FOREWORD FROM UN-HABITAT'S EXECUTIVE DIRECTOR

Artificial intelligence has already started to have an impact on urban settings at an unprecedented pace, with sophisticated solutions being deployed in the streets, at airports and in other city installations. In fact, cities are becoming experimental sites for new forms of artificial intelligence and automation technologies that are applied across a wide variety of sectors and places.

These developments, and emerging practices such as predictive policing, are dramatically changing cities and our societies at a time when the world is experiencing rapid urbanisation and a range of changes and challenges: climate change, the ongoing impact of the COVID-19 pandemic, access to basic urban services, infrastructure, housing, livelihoods, health and education. At the same time, AI and AI-enabled solutions are opening up new opportunities for cities while also being deemed to pose significant risks and challenges, such as potential bias and discrimination, privacy violations and other human rights violations, including surveillance schemes.

To support cities in their efforts to appropriately apply artificial intelligence, UN-Habitat has partnered with Mila – Quebec Artificial Intelligence Institute to provide reflections and guidance on AI and its responsible use in cities. The paper, which is part of our strategy to promote a people-centred approach to digital transformation, covers urban applications of AI, risks, sets out specific approaches and tools for urban AI governance, and provides a set of key recommendations for urban leaders implementing AI in local governments.

It is important for local (and national) governments to recognise the risks associated with the use of artificial intelligence that can arise as a result of flawed AI data, tools and recognition systems. Our preferred approach to AI in urban environments is anchored in the UN's Universal Declaration of Human Rights and UN-Habitat's specific mandate to promote inclusive, safe, resilient and sustainable cities (SDG 11). Our approach is aligned with our people-centred and climate-sensitive approaches to innovation and smart cities, which seek to make urban digital transformation work for the benefit of all, driving sustainability, inclusivity, prosperity and the realization of human rights in cities and human settlements. Our approach also focuses on addressing safety and urban planning concerns to ensure people-centred, safe and appropriate deployment of artificial intelligence within cities.

Within this approach, we emphasise the important role that governments, particularly local authorities, play in stewarding the necessary frameworks, infrastructure and capacity development to manage and govern the responsible deployment and use of AI-powered solutions.

Thanks to our partners at Mila for their collaboration.

I hope you find this report insightful and useful.

Ms. Maimunah Mohd Sharif
Under-Secretary-General and Executive Director
United Nations Human Settlements Programme
(UN-Habitat)

FOREWORD FROM VALÉRIE PISANO, PRESIDENT AND CEO OF MILA

Technological innovation, including artificial intelligence (AI), is (re)shaping how we approach almost every sphere of life. Urban environments are no exception to this transformation. AI systems can already be applied to key areas of urban intervention ranging from waste management, energy and transportation to public safety, healthcare and city governance. As AI continues to evolve, exciting opportunities that were once unimaginable will become available for cities and settlements to help them become more efficient and resilient in the face of today's challenges.

Like any other transformative opportunity, integrating AI into urban environments comes with challenges and risks that must be taken and tackled seriously for AI to benefit societies. Therefore, as we push the boundaries of AI integration in cities and settlements, these efforts must be rooted in and supportive of the human rights framework, as well as being sustainable, inclusive and aligned with local contexts. In other words, urban environments must adopt responsible AI technologies to succeed.

The development of socially responsible AI for the benefit of all is at the heart of Mila's mission. As a global leader in the field, Mila aims to contribute to the development of responsible AI and the fostering of social dialogue and engagement on this question, which is why we are proud to collaborate on this report with UN-Habitat. We hope this effort can support and inform civil society and public authorities as they navigate both the extraordinary benefits and the significant risks of AI-enabled technologies.

Through this interdisciplinary initiative that brings together experts in AI development, governance, public health, ethics, sustainability and urban development, we can start paving the way together for vibrant AI-powered cities that are climate-conscious, socially just and designed for all.

Valérie Pisano
President and CEO
Mila – Quebec Artificial Intelligence Institute

SECTION 1

Introduction

AI is a disruptive technology that offers a plethora of opportunities. This report presents an ambitious overview of some of the strategic applications of AI. Taking a risk-based approach, it also raises awareness of the risks of using AI, regardless of the application. The aim is to provide local authorities with the tools to assess where, and whether, AI could be valuable and appropriate, rather than instructing on what is or is not the right opportunity for a given context.

Cities and local authorities provide crucial areas for AI applications and policy-making because they regularly make day-to-day decisions about AI and how it affects people's lives. The emergence of AI technologies offers new ways to better manage and equip cities (UN-Habitat 2020, 180). However, the reshaping of cities through technology and innovation needs to reflect citizens' needs and, where possible, be used as a tool to foster more equal prosperity and sustainability. Cities, towns and settlements may have less policy and risk assessment capacity than nation states.

This report is part of UN-Habitat's strategy for guiding local authorities in supporting people-centred digital transformation processes in their cities or settlements. It is a collaboration with Mila–Quebec Artificial Intelligence Institute, a community of more than 1,000 researchers dedicated to scientific excellence and the development of responsible AI for the benefit of all. UN-Habitat helps build safe, resilient, inclusive and sustainable communities in over 90 countries by working directly with partners to respond to the UN Sustainable Development Goals (SDGs), and SDG 11 in particular. Together, this Mila–UN-Habitat collaboration offers a vision and understanding of how responsible AI systems could support the development of socially and environmentally sustainable cities and human settlements through knowledge, policy advice, technical assistance and collaborative action.

1.1 HOW TO READ THIS REPORT

The report is structured in five main parts: an introduction to AI, guidelines for AI governance, an overview of urban applications of AI, a risk assessment framework and a guide for AI strategy. Each section has a short overview summary.

While the report is lengthy, it is a tool for both policymakers and technical experts. This is a vertical document that works across an organisation. It can be used by local authorities, by city managers and directors, and by the technical people involved in the coding or maintenance of an AI system.

First, the report situates the discussion by describing AI: what it is, its different types, the opportunities it offers for cities, and its current limitations. The report then briefly discusses the importance of AI governance—which should be context-sensitive, anchored in and respectful of human rights, and centred on the public interest—as well as some of its key challenges.

The applications section follows; it identifies key sectors for intervention for cities, along with examples of AI applications within each of those pillars. Each application area is linked to the Sustainable Development Goal it supports, and a series of tags indicates high-impact, locally relevant and long-term endeavours.

The risk assessment framework presents an overview of the risks along the different phases of AI. There are many interrelations between risks. Each listed risk is accompanied by a set of reflective guiding questions. Defining success and evaluating the risks of an AI system should be done holistically, including both technical and contextual issues. Technical specifications alone will not determine an AI system's success; the social, political and structural contexts are crucial. The framework is designed for holistic reflection.

Finally, the report offers a guide of recommendations and areas of action to consider when building an AI strategy. While this guide is not exhaustive, its intention is to support local authorities by suggesting areas of action that can help ensure the development of responsible AI for cities and settlements.

1.2

GUIDING FRAMEWORKS

This report builds on existing frameworks that direct principles, values and policy actions in relation to artificial intelligence.

HUMAN RIGHTS

Human rights are the universal and inalienable rights of every human being, and they form the basis of all UN development approaches. They were institutionalised in the Universal Declaration of Human Rights by the United Nations General Assembly in 1948. The use of AI should be guided by these rights to ensure that no one is left behind, excluded or negatively impacted by its use.

SDGS

The UN's 17 Sustainable Development Goals (SDGs) are an integrated set of goals and targets for inclusive global development by 2030. SDG 11 calls for the safety, resilience, inclusivity and sustainability of cities and for enhancing participatory and integrated human settlement planning and management. The SDGs emphasise the importance of integrated approaches to using AI, considering a range of sectors and stakeholders.

NEW URBAN AGENDA

The New Urban Agenda (NUA) represents UN-Habitat's shared vision of how to achieve a more sustainable future for urbanisation (UN, 2017). It focuses on creating synergies across the mandates and strategic plans of different UN entities to maximise impact. The NUA encourages cities to develop frameworks for technologies such as AI to guide their development.

PEOPLE-CENTRED SMART CITIES

UN-Habitat's flagship programme, People-Centred Smart Cities, provides strategic and technical support to governments on digital transformation. It promotes the deployment of technological innovations to realise sustainability, inclusivity, prosperity and human rights and to make urban digital transformation work for the benefit of all. It leverages digital technologies for inclusive and sustainable development while preventing cities from having to constantly catch up (UN-Habitat, 2020a).

ETHICS OF AI

In November 2021, UNESCO adopted the first global instrument on the ethics of AI, which is strongly grounded in the recognition of the importance of promoting and protecting human rights. See the Recommendation on the Ethics of Artificial Intelligence (UNESCO, 2021).

SECTION 2

AI and Cities

2.1. WHAT IS AI?

Artificial intelligence as a technology is continuously expanding, so it is challenging to define AI as a whole. This continuous growth has led to a wide range of definitions of AI (Russell and Norvig, 2010; McCarthy, 2007). Colloquially, AI is often used as an umbrella term to cover a range of different types and technical subcategories.

At the core, AI is a system that produces outcomes based on a predefined objective (OECD, 2019). The objective for an AI is the translation of a human-defined goal into a mathematical one. Outcomes can be predictions, recommendations or decisions. For example, the human goal of winning a chess game can be translated into an objective of choosing a sequence of moves that maximises the probability of winning a chess game.

The terms **algorithm**, **AI system**, **AI ecosystem**, and **AI** indicate different scale levels. An **algorithm** is the most specific; it is a process which generates an output from an input. An **AI system** is usually a single application, using a unique input and producing its own output. An **AI ecosystem** is a network of AIs which interact with each other. **AI** is the broader term which includes all the different methods and systems, usually referred to the field as a whole.

While AI systems are complex, they tend to follow three major steps: engaging with data, abstracting the perceptions in the data, and formulating outcomes. Within the AI community, these steps are referred to as **the AI pipeline**.

2.1.1. WHAT IS RESPONSIBLE AI?

Digital innovation can be an inclusive force for good only if implemented with a firm commitment to improving people's lives and well being, as well as to building city systems that truly serve their communities (UN-Habitat, 2021, p. 6). However, good intentions alone are not sufficient to ensure AI systems are built responsibly. In fact, it is possible—and unfortunately not uncommon—even for an "AI for good" project or system to unintentionally replicate or compound social inequalities and biases. This is why responsible AI must be a central part of any discussion about AI development.

There is not yet consensus regarding a definition of responsible AI. In this report, it is referred to as an approach whereby the lifecycle of an AI system must be designed to uphold—if not enhance—a set of foundational values and principles, including the internationally agreed-upon human rights framework and SDGs, as well as ethical principles such as fairness, privacy and accountability. In this context, the objective of an AI system—for example, whether it aims to automate administrative tasks or to support the fight against the climate crisis ("AI for good")—is relevant yet secondary. Rather, responsible AI emphasises the importance of holistically and carefully thinking about the design of any AI system, regardless of its field of application or objective. It is therefore the collection of all choices—implicit and explicit—made in the design of the life cycle of an AI system that can make it either responsible or irresponsible. In summary, "ensuring responsible, ethical AI is more than designing systems whose result can be trusted. It is about the way we design them, why we design them, and who is involved in designing them" (Dignum, 2022).

This report is premised on the belief that AI applications should be without exception responsible in their design, and provides a general framework on how to deploy AI responsibly in the context of cities or settlements.

2.1.2. WHAT ARE DIFFERENT TYPES OF AI?

Understanding what AI is and isn't is important for non-technical users, because this knowledge allows them to reflect more critically about AI and enables them to participate in the necessary public conversations about its use, its governance, and ultimately, the place it should take within our societies, cities and settlements.

There are two main categories of AI systems: **Symbolic methods** rely on a series of predefined logical rules, and **statistical methods** identify patterns in a dataset to shape the outcomes (Clutton-Brock et al., 2021). For example, in a chess game, a symbolic AI system could choose moves based on a series of rules, such as "when having to choose between losing the queen or a pawn, sacrifice the pawn." In contrast, a statistical AI system can "learn" which moves are desirable based on a dataset of previous games.

In statistical methods, **machine learning** and **deep learning algorithms** are the most famous. There are three broad approaches within machine learning: supervised learning, unsupervised learning and reinforcement learning. The difference reflects the type of information the algorithm uses and whether or not it interacts with its environment. These differences are relevant because certain challenges arise out of this interaction.

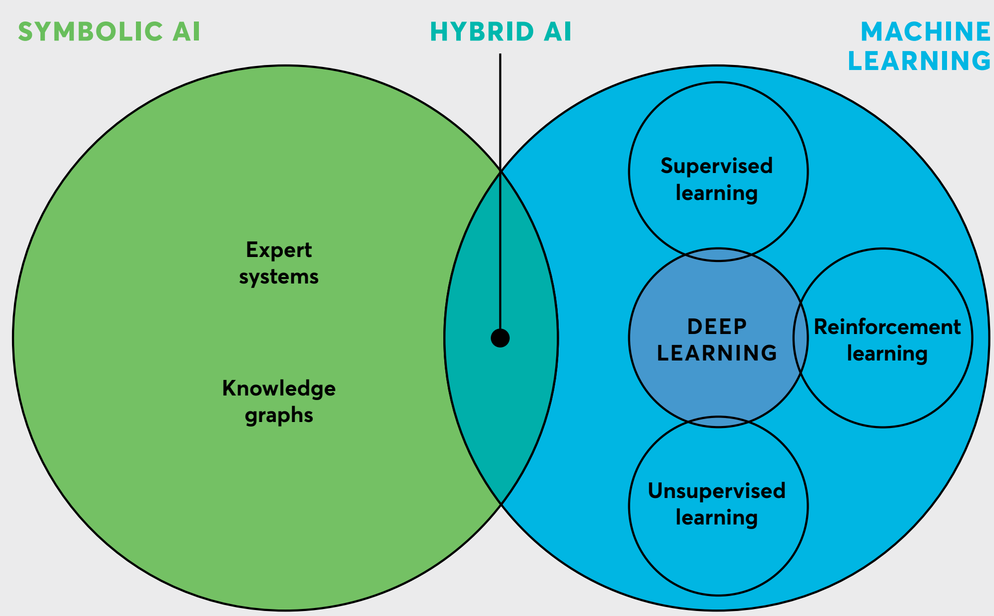
In **supervised learning**, the algorithm is trained with a labelled dataset of examples to learn a rule that can predict the label for a new input. For example, a training dataset may be images of pedestrian crossings that have been labelled for whether there is a person crossing or not. Because data labelling is usually done manually, it requires significant labour.

In **unsupervised learning**, algorithms use unlabelled datasets. These discover "hidden" structures in the dataset (e.g., clustering and visualisation (Murphy, 2012)) or are used as an extra step for other algorithms (e.g., representation learning (Bengio et al., 2013)).

In **reinforcement learning,** an algorithm interacts with its environment by choosing a sequence of actions to accomplish an objective. The goal is to learn the best sequence of actions through rewards or feedback from the environment. For example, autonomous navigation often uses reinforcement learning (Kiran et al., 2021); when the algorithm makes decisions about trajectory, safely controlling the vehicle provides larger rewards.

Deep learning algorithms are rapidly gaining popularity. According to Goodfellow et al. (2016, p. 5), "deep learning enables the computer to build complex concepts out of simpler concepts." This differs from traditional approaches to AI, which require large amounts of **feature engineering**, which uses human experts to extract attributes from raw data. Deep learning is compositional, combining a large hierarchy of learned concepts.

Artificial intelligence



2.1.3. DEEP LEARNING APPLICATIONS

Advancements in deep learning have led to significant innovations in AI applications. Many of these innovations power the types of AI cities will encounter and use.

For example, **speech recognition** is used in auto-transcription devices. In traditional machine learning, datasets used to contain labelled features such as volume or pitch. In contrast, in deep learning, an algorithm is provided with sound files and, in the training stage, conducts mathematical processing of the audio to infer its own attributes.

In the case of **image classification**, a deep learning algorithm aims to build higher-level graphic concepts such as "person" and "house" from lower-level concepts such as "corner" and "texture," and those are in turn constructed from "edges" or pixel values.

Computer vision has benefited greatly from deep learning innovations. For example, a computer vision task can have the objective of labelling the items in a set of images. In this way, when the algorithm is given a dataset of labelled images, it learns to distinguish between two items such as a cat and a car.

Lastly, **natural language processing** can be seen in chatbots. Chatbots can be trained to generate a conversation and automatically respond to people's text messages online. To do so, developers provide the chatbot with a dataset containing, for instance, all of the public written content on some social media platform.

2.1.4. OPPORTUNITIES FOR AI IN CITIES

Cities and settlements of all sizes and locations may benefit from using AI systems to address key urban challenges. Because AI is not specific to one domain or even to one technology, it has numerous applications for more sustainable and inclusive development. This report focuses on key areas of intervention for cities: energy, mobility, public safety, water and waste management, healthcare, urban planning and city governance.

Integrating AI systems could be a key to addressing social, economic and ecological challenges at a global scale. While every city is different, cities are the centre of societal transformations, and digital transformation in particular. They are where people, jobs, research, wealth and leisure concentrate. They concentrate access to opportunities for a greater number of people, as well as concentrating societal and environmental issues. Because cities play a role in global networks, the benefits and the risks of integrating AI systems also extend well beyond their borders.

As cities continue to experience significant challenges relating to resource demands, governance complexity, socioeconomic inequality and environmental threats, innovation is necessary for tackling emerging problems (Yigitcanlar et al., 2021). To take full advantage of the potential of AI for cities, local governments can create the enabling conditions for sustainable and inclusive development. Guiding the development of these conditions, together with a careful balancing of the opportunities and risks, is the purpose of AI governance.

2.1.5. LIMITATIONS OF AI

Applying AI responsibly requires understanding the key limitations of AI systems. Three in particular will always be present:

AI systems reinforce the assumptions in their data and design. In order for an algorithm to reason, it must gain an understanding of its environment. This understanding is provided by the data. Whatever assumptions and biases are represented in the dataset will be reproduced in how the algorithm reasons and what output it produces. Similarly, design choices are made all along the AI life cycle, and each of these decisions affects the way an algorithm functions. Because negative societal assumptions may be reflected in the dataset and design choices, algorithms are not immune to the discriminatory biases embedded in society.

> Algorithms can reproduce gender stereotypes based on their dataset. Chatbots, for example, include algorithms trained on large datasets of text to learn existing relations between words (Garg et al., 2018). The result is that women are associated with words such as "nurse" and "receptionist," while men are associated with the words "doctor" and "engineer" (Bolukbasi et al., 2016).

AI systems cannot evaluate their own performance. AI systems measure their performance against pre-defined optimisation goals, but these are isolated from the greater context. The term *artificial intelligence* is somewhat of a misnomer; human reasoning can judge the relevance of a type of knowledge to a specific situation, but algorithms cannot. While it may be tempting to see algorithms as neutral "thinkers," they are neither neutral nor thinkers. This means that an AI system's objective must be carefully aligned with human goals, and considerable attention is required for monitoring and evaluating by humans.

AI systems are mathematical and cannot integrate nuance. Defining an AI system's objective requires translating a human goal into a mathematical formula. This creates strict constraints on what types of knowledge and information can be integrated into an algorithm's reasoning. Because everything needs to be concretely defined, algorithms are unable to comprehend a whole range of subjective, qualitative and nuanced information.

2.2. GOVERNANCE

2.2.1. GOVERNANCE OVERVIEW

AI is not neutral, and context matters. Understanding how structures embed values offers local authorities the possibility to direct the development of AI towards key values for inclusive and sustainable development. This requires an understanding of AI governance, which is the sum of AI regulations, ethics, norms, administrative procedures and social processes.

Readers should keep these issues in mind while reading the rest of this report in order to better anchor the applications and risks in the local context and in local definitions of the public interest.

This section is short. It presents an overview of what AI governance is, why it matters and what challenges are particular to AI in an urban setting. Section 5 then makes specific recommendations on how to design and implement a strategy.

2.2.2. WHY GOVERNANCE MATTERS

Governance is a tool to direct the development of AI towards a set of values, such as inclusive and sustainable development. Directing urban development towards the public interest and respect for human rights is a conscious choice. If this choice is not made, the structures and processes of AI and its governance will embed values unconsciously, causing significant risks (see section 4.2.1). **AI is not neutral**; both formal structures, such as the design of algorithms, and informal arrangements, such as social norms, embed and propagate values.

The **context matters** for AI adoption. The technical choices around developing AI are important, and much of this report outlines these. However, the success of an AI system is often dependent on what happens once the algorithm moves out of the laboratory and into the real world, where there are people (see section 4.2.4). AI technologies are co-created within a society, especially when AI uses citizens' data and shapes their lives. Each urban settlement will have its own unique context, with its own interrelationship of social norms, values and ways of working. Governance is a tool that local authorities can use to balance opportunities and risks of AI in a manner most appropriate to their local context.

AI governance, like digital governance more broadly, combines regulations, ethics, norms and social practices. Governance is greater than the sum of its parts; it also includes the process of how to make decisions about these aspects and the social relations that shape these decisions (Floridi, 2018; Jameson et al., 2021). This definition is intentionally broad, as the meaning of "governance" can vary depending on the discipline and context in which it is used. For example, many definitions focus only on administrative rules and tools to fulfil legal and ethical requirements (Mäntymäki et al., 2022). For a city directing inclusive urban development, self-regulation and ethics processes must be complemented with building capacities and hard laws (Larsson and Heintz, 2020; Ala-Pietilä and Smuha, 2021; Wirtz et al., 2020).

2.2.3.

AI IN THE CITY

AI affects the city, and it changes the environment that local authorities are working in. As AI systems evolve to change urban socioeconomic development, they will also shift social organisation in the city. For example, a major dynamic in cities is platform urbanism, where the increasing use of platforms, with their predictive tools and need for data, reshapes relationships of labour, mobility, consumption and governance (Leszczynski, 2020).

These are systemic issues. As a result, implementing AI in a city combines with and compounds existing dynamics so that new systemic issues emerge. For example, AI can have adverse consequences on city residents' rights by disproportionately shifting power dynamics in a negative direction or maintaining existing negative power dynamics (Rodriguez, 2021). An AI system may also intensify inequality between those who have access to its services and those who don't, or between those who benefit from it and those who don't (see the "digital divide" subsection under section 4.2.2.1). The combination of different data sources and predictive tools creates new knowledge about city residents, which in certain contexts may support existing instances of rights abuses (Reuter, 2020; Aitken, 2017).

To address this changing environment, local authorities are in a unique position to shape the city's context. They can create an enabling environment for the development of AI that in turn enables sustainable and inclusive development. Through the regulatory environment and with each contract awarded, local authorities shape what services are implemented in the city, thus creating the conditions for future development. They can set conditions for investment in technology and infrastructure, enable a bustling civil society, and foster an innovative environment to advance the public interest. In order for the city to be proactive rather than reactive, digital innovation requires a clear city-level strategy. Specific recommendations for how to create such a strategy are presented in the section 5.

2.2.4.

KEY CHALLENGES FOR AI GOVERNANCE

MULTI-LEVEL GOVERNANCE

While many countries have already released national guidelines on AI (Schmitt, 2021), local governments still face various challenges regarding the development, implementation and evaluation of regulatory frameworks for policy capacity (Taeihagh, 2021). A city is itself an actor within a larger territorial entity, an independent system within larger systems, and a node within a global network of cities. The nested geographic aspect of a city's territory is reflected in the dispersion of governance capacities and responsibilities across multiple jurisdictions. The different levels of technological governance involve interdependent actors from all levels of government operating on the city's territory, as well as non-state actors, including the public, private and social sectors (Enderlein et al., 2010).

The resulting multi-level organisation poses a genuine challenge as it may imply limited capacity and resources for local authorities. However, multi-level governance approaches may support cities in benefiting from these jurisdictional structures by attributing each level of governance with a key role in decision-making.

The role of national governments in particular is to consolidate specific principles, policy frames and values that are active in national discourses on AI governance and to cultivate a national ecosystem for AI innovation and development. While cities will refer to these national guidelines when they exist, cities can still actively participate in establishing priorities and bringing forward their interests and their vision of AI (UN-Habitat, 2021).

ACCOUNTABILITY

AI comes with three major accountability challenges: political accountability, adapting to changes over time, and the responsibilities of automated systems. Overall, accountability shapes much of how the city balances risks and opportunities in developing urban AI. This makes accountability structures especially important to consider in AI governance. Here, we focus on the latter two challenges.

It's incredibly important to consider who is accountable after an AI system is delivered or procured. Algorithmic systems will change over time, and their impacts are not always predictable. AI systems will shape and be shaped by the environment in which they are deployed, and a change of purpose over time may challenge agreements that were made in the early stages of the AI life cycle. One example is mission creep, a relatively common occurrence when technologies are intentionally repurposed for surveillance practices (see section 4.2.1).

There are also ever-increasing concerns regarding the responsibilities of automated systems. Algorithms act on the world without transparent human intention, which challenges our existing human-centred accountability frameworks. While autonomous systems may act relatively independently, they are still designed, funded and owned by human actors. As a result, allocating responsibility among the actors designing, funding or, in some cases, using the algorithm is important. For instance, if a self-driving car crashes, who is liable?

The design and implementation of an algorithm in a city's critical services shifts existing responsibilities amongst actors. For instance, a public-private partnership created for public transport services may rely on a multitude of subcontractors that interact with AI systems and process sensitive data. The accountability challenge is to acknowledge responsibility throughout an AI system's life cycle and create governance mechanisms to respond to issues if and when something goes wrong.

LIMITED CAPACITY

Limited capacity is one of the biggest challenges for AI governance in cities across the world. While the lack of AI skills among professionals may be exacerbated by a city's socio-economic situation and the global digital divide, it is a challenge for rich and poor cities alike.

The global move towards AI has spurred an increasing demand for IT professionals. In the race for IT talent and more niche AI skills, the private sector outpaces governments in their ability to attract these professionals. Notably, cities lack sufficient funds to hire specialised human resources and drive cutting-edge development work in-house. This limited capacity puts cities in a position where the technical expertise needed for the proper governance of AI is almost always outsourced and procured.

Furthermore, the skill gap between the decision-makers who are responsible for funding AI solutions and those who will provide the technology renders system monitoring very difficult. Through cross-sectoral collaborations (short to medium term) and local capacity-building (long term), the city may overcome this limitation while centring its values. These two levers form a significant part of the following governance recommendations (see section 5.3).

SECTION 3

Applications

3.1.

Applications overview

AI is evolving at such a rapid pace that the potential number of applications in the urban context is now tremendous. In order to make this conversation digestible, this section identifies key sectors for intervention for cities, along with examples of AI applications within each of those pillars.

The key sectors outlined include energy, mobility, public safety, water and waste management, healthcare, urban planning and city governance. This outline is not exhaustive.

Each application presented is a concrete example of an existing technology, not of futuristic innovations. The examples also explicitly support sustainable, low-carbon, inclusive development.

Each application area is linked to the Sustainable Development Goal it supports. A series of tags is used to indicate high-impact, locally relevant and long-term endeavours.

Lastly, the applications are linked to the Risk Framework which follows. Particular applications are linked to specific risks from the Risk Framework where appropriate, though these links are for illustrative purposes. For a full view, readers are encouraged to refer to the section 4.

3.2.

Energy

SDGs: 7.1, 7.2, 7.3, 7.a, 7.b, 9.1, 9.4, 9.5, 9.b, 11.b, 13.2

RISK: Mission creep

AI can be a useful tool across electricity systems and in accelerating the transition to a low-carbon society. Indeed, AI can both reduce emissions from existing power plants and enable the transition to carbon-free systems, while also improving the efficiency of systems. In particular, AI can contribute to power generation by forecasting supply from variable sources and through predictive maintenance. Working in this area requires close collaborations with electricity system decision-makers and practitioners in fields such as electrical engineering.

RISK: Mission creep

Policymakers and stakeholders considering applications of AI in power generation must be mindful not to impede or delay the transition to a low-carbon electricity system. For example, using AI to prolong the usable lifetime of a high-emissions coal power plant, or to accelerate the extraction of fossil fuels, could run counter to climate and air quality goals. Ideally, projects should be preceded by system impact analyses that consider effects on society and the environment. Such early impact assessment can ensure that projects do not enable or perpetuate unsustainable behaviours.

3.2.1.

FORECASTING ELECTRICITY GENERATION

With the transition to renewable energy, electricity generation will become both more intermittent and better distributed. This is because the output of such generation will be determined by local environmental conditions (e.g., wind speeds and cloud cover in the case of wind turbines and solar panels, respectively) that can vary significantly. Importantly, AI models can employ various types of data, such as satellite images and video feeds, to create forecasts to understand the emissions from different sources. Here, AI offers additional opportunities by:

- Enabling forecasts of wind- and weather-generated power quantities by analysing patterns in historical data. These can provide much-needed foresight in contexts such as power system optimisation, infrastructure planning and disaster management (Mathe et al., 2019; Das et al., 2018; Voyant et al., 2017; Wan et al., 2015; Foley et al., 2012).
- AI predictions have been used as information and evidence to help determine the location and need for new plants, supporting operators and investors.

Forecasting with hybrid physical models

Many AI techniques are "domain-agnostic," meaning they can easily be applied to different domains. In an optimal scenario, however, the AIs of the future will improve predictions by incorporating domain-specific insights. This is particularly important for forecasting electricity generation: since weather drives both electricity generation and demand, AI algorithms forecasting these quantities should not disregard established techniques for climate and weather forecasting. Such hybrid models involving both physics and AI can not only help with getting more reliable forecasts in the long term, but also help with the prediction of catastrophic weather events (see section 3.4).

3.2.2. PREDICTIVE MAINTENANCE OF EXISTING INFRASTRUCTURE

Optimising maintenance on existing power generation systems has multiple advantages over building new infrastructure, as it can help minimise emissions and reduce the need for costly financial investments in new infrastructure (Sun and You, 2021). In this context, AI has been successfully used to operate the diagnostics and maintenance of existing systems through sensor data and satellite imagery. In particular, AI has helped with:

- Detecting leaks in natural gas pipelines (Wan et al., 2011; Southwest Research Institute, 2016)
- Detecting faults in rooftop solar panels (Bhattarcharya and Sinha, 2017)
- Detecting cracks and anomalies from image and video data (Nguyen et al., 2018)
- Preemptively identifying faults from sensors and simulation data (Caliva et al., 2018).

3.2.3. ACCELERATING EXPERIMENTATION

AI can be used to accelerate scientific discovery in areas such as materials sciences. In such cases, AI is not a replacement for scientific experimentation, but it can learn from past experiments in order to suggest future experiments that are more likely to be successful (Clutton-Brock et al., 2021).

- AI is being used to accelerate the development of new materials that can better store and harness energy from variable low-carbon sources (e.g., batteries and photovoltaic cells) (Butler et al., 2018; Liu et al., 2017; Gómez-Bombarelli et al., 2018).
- AI has also been used to understand innovation processes in order to inform policy for accelerating material science (Venugopalan and Rai, 2015).

3.2.4.

TRANSMISSION AND DISTRIBUTION

SDGs: 7.1, 7.2, 7.3, 7.a, 7.b, 11.a, 11.b, 13.2

Power grids require balance between the supply and demand of energy. This balance can be affected by multiple factors, including unexpected fluctuations in supply or demand, the algorithms used to control grid infrastructure, and failures or weaknesses in that infrastructure. In most countries, the electricity grid has changed very little since it was first installed. Moreover, existing grids were designed based on the idea that electricity is produced by a relatively small number of large power stations that burn fossil fuels and is delivered to a much larger number of customers, often some distance from these generators, on demand (Ramchurn et al., 2012). Aggravating this, the grid itself relies on aging infrastructure plagued by poor information flow (for example, most domestic electricity meters are read at intervals of several months) and has significant inefficiencies arising from loss of electricity within the transmission networks (on a national level) and distribution networks (on a local level).

3.2.5.

SYSTEM OPTIMISATION AND CONTROL

HIGH IMPACT

LOCALLY RELEVANT

RISKS: Explainability, privacy and privacy attacks

When balancing electricity systems, operators need to determine how much power every controllable generator should produce in a process called scheduling and dispatch. With the aim to achieve optimal power flow, this process must also be coordinated at different time scales. The balancing process becomes even more complex as electricity systems include more storage, variable energy sources and flexible demands. Traditional power system monitoring, optimisation and intelligent control methods are proving inadequate for these purposes, pushing cities towards the use of AI to optimise and balance power grids in real time (smart grids) (Ramchurn et al., 2012; Perera et al., 2014; Victor, 2019).

- In this context, AI has been used to improve scheduling and dispatch processes by improving the quality of flow optimisation solutions (Borgs et al., 2014; Dobbe et al., 2017; Dobbe et al., 2018).
- AI is also used to learn from the actions of human engineers working with power-system control (Donnot et al., 2017).
- AI techniques have been used to ensure that the distribution system runs smoothly by estimating the state of the system even when only few sensors are available (Donti et al., 2018; Jiang et Zhang, 2016; Pertl et al., 2016).
- Image processing, clustering and optimisation techniques have also been used on satellite imagery to inform electrification initiatives (Ellman, 2015). Traditionally, figuring out what clean electrification methods are best for different areas can require slow and intensive survey work, but AI can help scale this work.

AI models can also help operate rural microgrids—that is to say, localised, self-sufficient energy grids—through accurate forecasts of demand and power production, since small microgrids are even harder to balance than country-scale electric grids (Cenek et al., 2018; Otieno et al., 2018). As these new local sources of energy are also emerging, decentralised energy generation will be increasingly important to future energy systems and will actively contribute to the optimisation of the grid. For example, homes equipped with smart meters and fitted with clean energy sources, as well as newly developed energy storage, can plug into the grid and supply energy into distributed energy networks.

AI can also help integrate rooftop solar panels into the electric grid (Malof et al., 2016; Yu et al., 2018). In the United States and Europe, for instance, rooftop solar panels are connected to a part of the electric grid called the distribution grid, which traditionally did not have many sensors because it was only used to deliver electricity one way, from centralised power plants to consumers. However, rooftop solar and other distributed energy resources have created a two-way flow of electricity on distribution grids.

Definition: Microgrids

A microgrid is a self-sufficient energy system that serves a discrete geographic area, such as a college campus, a hospital complex or a neighbourhood. Within a microgrid are one or more kinds of distributed energy (solar panels, wind turbines, combined heat and power, generators) that produce its power. Microgrids are an important step towards energy equity. They are uniquely suited to help empower disadvantaged communities. This is because of their robustness; they can provide energy both with the centralised grid and independently from it, providing power even when the centralised grid is unavailable.

RISK: Explainability

When AI helps control electrical grids, system developers may require technical details about how algorithms function, regulators may require assurance about how data is processed, and households may want their smart meters to provide accessible information through an intuitive user interface. Explainability should be carefully considered: the more critical a system is and the greater the cost of failure, the more accurate an explanation is required. While the risk of an insufficient explanation is low in the case of an app helping a user understand household energy consumption, it is much greater when it comes to digital interfaces informing grid operators. Indeed, incorrect interpretation of outputs and model usage could lead to grid failure, which is often a catastrophic event.

RISK: Privacy and privacy attacks

Relevant stakeholders should be aware that sharing data about critical infrastructure such as energy systems without adequate protection may pose a risk to cybersecurity and system resilience.

3.2.6. FORECASTING SUPPLY AND DEMAND

Since energy generation and demand both fluctuate, real-time electricity scheduling and longer-term system planning require forecasting ahead of time. Better short-term forecasts can also reduce the reliance on standby plants, which often rely on fossil fuels, and help proactively manage increasing amounts of low-carbon variable generation. In this context, hybrid AI physical models can contribute to modelling the availability of different energy sources and to forecasting supply and demand across the system (see section 3.2.1).

- Hybrid AI-physical models have been used for modelling precise energy demand within buildings (Robinson et al., 2017).
- Through hybrid AI-physical models, it has also been possible to model energy dynamics in an urban microclimate, such as for instance a campus or a neighbourhood (Nutkiewicz et al., 2018).
- AI has also been used to understand specific categories of demand, for instance by clustering households into groups with similar patterns of energy use (Zhang et al., 2018).

3.2.7.

PREDICTIVE MAINTENANCE FOR THE TRANSMISSION AND DISTRIBUTION NETWORK

RISK: Digital divides

Predictive maintenance is a key strategy for AI to contribute to decreasing emissions while increasing infrastructure safety, driving down costs and increasing energy efficiency. During transmission and distribution, predictive maintenance can help prevent avoidable losses.

- AI can suggest proactive electricity grid upgrades. AI can analyse the grid information at any given time and determine the health of the grid, avoiding wastage of generated power (see sources cited in section 3.2.2).
- Similar techniques also help reduce the amount of energy consumed during transmission (Muhammad et al., 2019).

While some of these losses are unavoidable, others can be significantly mitigated to reduce waste and emissions.

3.2.8.

ENERGY USE AND EFFICIENCY

SDGs: 7.1, 7.2, 7.3, 7.a, 7.b, 11.6, 11.a, 11.b, 13.2

3.2.8.1.

Forecasting energy use and improving efficiency

Following the transition to variable low-carbon energy, the supply and price of electricity will vary over time. Thus, energy flexibility in buildings will be required in order to schedule consumption when supply is high. For this, automated demand-side forecasts can respond to electricity prices, smart meter signals or learned user preferences and help efficiently schedule energy use.

RISKS: Explainability, privacy and privacy attacks

AI can enable more flexible industrial electrical loading by optimising a firm's demand response to electricity prices. There are several promising ways to enhance the operating performance of heavy-consumption energy systems using AI, as for instance:

- Demand response optimisation algorithms can help adjust the timing of energy-intensive processes such as cement crushing and powder coating to take advantage of electricity price fluctuations (Zhang et al., 2016).
- AI is also used to make sense of the data produced by meters and home energy monitors.

3.2.8.2.

Controlling energy usage

LOCALLY RELEVANT

RISKS: Explainability, privacy and privacy attacks, sustainability

There is significant scope for AI to help in reducing energy use and increasing efficiency in industrial, residential and commercial settings. Such applications can both reduce energy bills and lower associated carbon emissions.

- In this context, AI can be employed to analyse real-time building data and then provide insights on building performance.
- As building-related applications must transmit high volumes of data in real time, AI is also key to pre-processing large amounts of data in large sensor networks, allowing only what is relevant to be transmitted instead of all the raw data that is being collected.
- AI can be used to forecast what temperatures are needed throughout the system, better control to achieve those temperatures, and provide fault detection.
- AI can also be employed to adjust how many systems within buildings, such as lights and heating, operate based on whether a building or room is occupied, thereby improving both occupant comfort and energy use.

Definition: Remote sensing

RISK: Sustainability

Remote sensing is the process of detecting and monitoring the physical characteristics of an area by measuring the radiation reflected and emitted at a distance (typically from cameras on satellites or aircrafts). Images are collected remotely from either normal or specialised cameras (such as infrared ones), helping researchers "sense" things about an object or a phenomenon.

RISK: Sustainability, mission creep and privacy violations

In the future, sensors could be everywhere, monitoring every source of emissions as well as every person's livelihood. While this might start from a good intention, it could go too far. Moreover, the swarm of new sensing devices could add layers of embodied emissions and material waste into the environment, and the processing and storage of the vast amount of data they generate might come with a cost in emissions too. Finally, data collected through remote sensing could end up being maliciously used for surveillance and privacy breaching applications. (For the risks relating to these applications, see the "AI in surveillance" box in section 3.4.)

Definition: Digital twins

Data obtained from remote sensing can be used to model a digital twin. A digital twin is a digital representation of a physical object, process or service. A digital twin can be a digital replica of an object in the physical world, such as a jet engine or a wind farm, or even larger items such as buildings. A digital twin is, in essence, a computer program that uses real-world data to create simulations that can predict how a product or process will perform. While digital twins are a powerful tool to simulate contained systems such as industrial processes, they are not equally well suited to simulating social phenomena. For a more in-depth discussion on how digital twins might not be the best solution when modelling social processes, see Takahashi (2020).

3.3.

Mobility

3.3.1.

PUBLIC TRANSPORT

SDGs: 11.2, 11.3, 11.6

3.3.1.1.

Transportation demand and forecasting

HIGH IMPACT

RISK: Privacy

AI can improve estimates of transportation usage, as well as modelling demand for public transportation and infrastructure. In turn, modelling demand can help in planning new infrastructure. Data-enabled mobility platforms can enable users to access, pay for and retrieve real-time information on a range of transport options. This can promote public transport usage and make it easier for individuals to complete their journeys. In this context, AI is particularly apt for processing information from diverse sources of data:

- AI is being used to learn about public transport usage through smart-card data. Moreover, AI modelling in combination with mobile phone sensor data can provide new means to understand personal travel demand and the urban topology, such as walking route choices (Manley et al., 2018; Ghaemi et al., 2017; Tribby et al., 2017).
- AI has been used to improve the short-term forecasting of public transit ridership (Dai et al., 2018; Noursalehi et al., 2018).
- AI has also been used to reveal the preferences of customers traveling by high-speed rail (Sun et al., 2018).
- AI can be applied to make public transportation faster and easier to use. For example, AI methods have been used to predict bus arrival times and their uncertainty (Mazloumi et al., 2011).
- Similarly, AI can improve the operational efficiency of aviation, predicting runway demand and aircraft taxi time in order to reduce the excess fuel burned in the air and on the ground due to congestion in airports (Jacquillat and Odoni, 2018; Lee, Malik et al., 2015).

Case study: City of Los Angeles Mobility Data Specification

Data is often proprietary. To obtain this data, the city of Los Angeles now requires all mobility-as-a-service providers, i.e., vehicle-sharing companies, to use an open-source API (Application Programming Interface). This way, data on the location, use and condition of all those vehicles are transmitted to the city, which can then use that data for their own services as well as guiding regulation (Open Mobility Foundation, 2021).

3.3.1.2.

Facilitating ride-sharing

Recent years have seen the advent of “as a service” business models, where companies provide a service rather than selling a product or commodity. In mobility, “mobility as a service” refers to a shift away from personally owned modes of transport towards a mobility service provided to a pool of users. In many cities, companies offer people a smartphone app to locate and rent a bike for a short period. Mobility as a service may also potentially reduce congestion and greenhouse gas emissions by increasing public transport.

FORECASTING

AI can be used for predicting demand in ride-sharing systems. Data on mobility as a service may also be useful for municipalities, for example in helping to understand how a rideshare service affects urban geography and transport patterns. In cases where such services are run by private entities, data-sharing agreements may be valuable to municipal entities.

SYSTEM OPTIMISATION

AI techniques make it possible to optimise the use of existing physical ride-sharing infrastructure in multiple ways. Business and operating models that offer mobility as a service can leverage digital technologies to support more fundamental changes to how individuals access transport services, reducing the number of personal vehicles on the road.

- Applying AI to survey data has helped relevant stake-holders understand public opinion on ride-sharing systems, such as, for instance, dockless bike-sharing (Rahim Taleqani et al., 2019).
- One challenge is the bike-sharing rebalancing problem, where shared bikes accumulate in one location and are lacking in other locations. In this context, AI can help by improving forecasts of bike demand and inventory. Moreover, AI can help to understand how usage patterns for bike stations depend on their immediate urban surroundings (Regue and Recker, 2014).

OPTIMISING INTERMODALITY

When traveling by train, the trip to and from the station will often be by car, taxi, bus or bike. There are many opportunities for AI to facilitate a better integration of modes in both the passenger and freight sectors. As an example, bike-sharing and electric scooter services can offer alternatives for urban mobility that do not require ownership and integrate well with public transportation.

AI has been used to help integrate bike shares with other modes of transportation by producing accurate travel estimates (Ghanem et al., 2017).

In the freight sector, AI has been applied to analyse modal trade-offs, that is, exchanges of shipments from different types of modalities (e.g., boat to truck, truck to delivery van) (Samimi et al., 2011).

Case study: Citymapper

Making transportation data openly available has supported the development of apps such as Citymapper, which provides access to information about traffic in real time so individuals can choose less congested routes, thereby improving individ-ual journeys as well as reducing congestion and emissions.

3.3.2.

PRIVATE TRANSPORT

SDGs: 7.3, 7.a, 9.b

3.3.2.1.

Improving electric vehicle technology

LONG-TERM ENDEAVOUR

To advance the adoption of electric vehicles (EV), AI is a decisive technology for EV costs and usability. Work in this area has focused on predicting battery state, degradation and remaining lifetime. With the aim of accelerating the technology behind EVs:

- AI can optimise battery charging by suggesting users where to position themselves during wireless charging and help provide users with better battery charge estimation (Hansen and Wang, 2005; Tavakoli and Pantic, 2017).
- Battery electric vehicles are typically not used for more than a fraction of the day, potentially allowing them to act as energy storage for the grid at other times. AI can help with optimising when energy should be transmitted from vehicle to grid or vice versa.
- AI can also inform the design of batteries and next-generation fuels to optimise the energy consumption of EVs (Fujimura et al., 2013).

3.3.2.2.

Facilitating autonomous vehicles' safety and adoption

RISKS: Societal harm, distress in the local labour market

AI is essential to many aspects of the development of autonomous vehicles (AVs).

- AI is involved in many tasks at the core of autonomous vehicles' functioning, including following the road, detecting obstacles, interpreting other vehicles' trajectories, managing speed, understanding driving styles and more.
- Further, AI can help to develop AV technologies specifically aimed at reducing congestion as well as fuel consumption. For example, AV controllers can be used to smooth out traffic involving non-autonomous vehicles, reducing congestion-related energy consumption (Wu et al., 2017; Wu, Raghavendra et al., 2022).

RISK: Societal harm

While autonomous buses could decrease greenhouse gas emissions, self-driving personal vehicles, on the other hand, may increase emissions by making driving easier, as well as attracting people towards private vehicle ownership and thereby augmenting the industrial production of vehicles. In the long run, this could cause harm to the environment and to society in general.

RISK: Distress in the local labour market

Vehicle automation could lead to the replacement of multiple types of human resources, such as bus, train and truck drivers. Even in situations in which automated vehicles do not explicitly replace human workers, their functions could be reduced to precarious unpaid or low-paid labour.

3.3.3. TRANSPORTATION INFRASTRUCTURE

SDGs: 11.2, 11.3, 11.6

3.3.3.1. Optimising electric vehicle infrastructure

In the context of building the appropriate infrastructure for electric vehicles to coexist with traditional mobility options within the city, AI can help in multiple ways.

- In-vehicle sensors and communication data are increasingly becoming available and offer an opportunity to understand the travel and charging behaviour of EV owners, which can, for example, inform the placement of charging stations. AI can also inform the positioning of battery charging towers in the city to facilitate usage and consumer adoption of EVs (Tao et al., 2018).
- Moreover, AI can help modelling EV users' charging behaviour, which in turn can inform the positioning of battery charging towers. This will be equally useful for grid operators looking to predict electric load (see section 3.2) (Wang, Li et al., 2019).

3.3.3.2. Optimising traffic flow and control

HIGH IMPACT

AI can be used for both vehicular and pedestrian traffic forecasting based on data obtained from dedicated sensors, such as traffic cameras, and from soft-sensing, such as mobile devices. Moreover, AI is being applied to understand how vehicles are moving around city centres, and in places has helped improve congestion prediction by changing street design and controlling traffic lights. The information can be used to ease traffic as well as reduce emissions. Traditionally, traffic is monitored with ground-based counters that are installed on selected roads and sometimes with video systems, in particular when counting pedestrians and cyclists.

- AI can help with this by automating traffic monitoring through computer vision.
- AI methods have made it easier to classify roads with similar traffic patterns. In this context, remote sensing is key to inferring infrastructure data, as satellite data present a source of information that is globally available and largely consistent worldwide (Krile et al., 2015; Tsapakis and Schneider, 2015; Gastaldi et al., 2013).
- As vehicles can be detected from high-resolution satellite images with high accuracy, AI and vehicle image counts can serve to estimate average vehicle traffic (Kaack et al., 2019).
- AI systems have been used for traffic light (signal) control.
- AI has been used to create platforms for interactive data manipulation to monitor and predict traffic behaviour, while potentially testing out planning scenarios at the same time.

- AI has also been used to prevent the escalation of traffic problems by finding mechanisms for fleet operators and cities to work together, for example by sharing data about congestion or pollution hotspots and rerouting around the problem before it becomes serious.
- The same sensors used for traffic prediction can also be used by an AI to determine how many pedestrians are waiting at the light and how much time it might require them to cross a street (Zaki and Sayed, 2016).
- Smart parking through AI has also been piloted, deploying sensors in parking spaces and communicating the information to road users through apps, with the potential to halve congestion in busy city areas (BT, n. d.).

AI can provide information about mobility patterns, which is directly necessary for agent-based travel demand models, one of the main transport planning tools.

- For example, AI makes it possible to estimate origin-destination demand from traffic counts, and it offers new methods for spatio-temporal road traffic forecasting.
- AI has been used to improve our understanding about passengers' travel mode choices, which in turn informs transportation planning, such as where public transit should be built (Omrani, 2015; Nam et al., 2017; Hagenauer and Helbich, 2017).
- Using AI on survey data can also help with understanding passengers' reasons for choosing a certain mode of transport (Seo et al., 2019).

3.3.3.3.

Predictive maintenance for roads and rails

- In road networks, it is possible to incorporate flood hazard and traffic information in order to uncover vulnerable stretches of road, especially those with few alternative routes. If traffic data are not directly available, it is possible to construct proxies from mobile phone usage and city-wide CCTV streams; these are promising in rapidly developing urban centres.
- AI can help to improve and optimise transportation infrastructure, for example by reducing the operations and maintenance costs of road surface quality and rail. Tools for efficiently managing limited resources for maintenance include predictive maintenance and anomaly detection. In predictive maintenance, operations are prioritised according to the predicted probability of a near-term breakdown.
- Remote sensing can be used to predict road and track degradation (Soleimanmeigouni et al., 2018).
- For anomaly detection, failures are discovered as soon as they occur, without having to wait for inspectors to show up or complaints to stream in.

3.4.

Public safety

RISKS: Explainability (see section 3.2), mission creep, societal harm

Cities are prime victims of many disastrous events, from extreme storms to earthquakes, and, as a result, they are at the forefront of disaster management. This section reviews various applications through which AI can help mitigate disasters, aid relief and support affected populations. For more ways in which AI applications can effectively support and complement urban governance in preventing disasters, please see section 3.7.

3.4.1.

ENVIRONMENTAL SAFETY

3.4.1.1.

Extreme weather event forecasting

AI can predict localised flooding patterns from past data, which could inform individuals buying insurance or homes. Since AI systems are effective at predicting local flooding during extreme weather events, these could be used to update local flood risk estimates to benefit individuals.

DANGER: AI in policing

Around the world, many law enforcement agencies have turned to AI as a tool for detecting and prosecuting crimes (Almeida et al., 2021). AI applications for policing include both predictive policing tools (as in the use of AI to identify potential criminal activities) and facial recognition technology. All these technologies have been shown to be biased in multiple ways and to lead to harsher impacts on vulnerable communities. For example, the COMPAS algorithm, used to predict the likely recidivism rate of a defendant, was twice as likely to classify Black defendants as being at a higher risk of recidivism than they were while predicting white defendants to be less risky than they were (Larson et al., 2016). By using the past to predict the future, predictive policing tools reproduce discriminatory patterns and often result in negative feedback loops, leading the police to focus on the same neighbourhoods repeatedly, and therefore leading to more arrests in those neighbourhoods. For example, the Strategic Subject List algorithm used data from previous police records to predict how likely people were to be involved in violence, without making any distinction between the victims and the perpetrators (Asaro, 2019). The Chicago police department then used the algorithm to create a "heat list," using it as a suspect list and surveillance tool; the people on it were therefore more likely to be arrested and detained (Asaro, 2019). Similarly, facial recognition technology that shows poor accuracy for certain demographics has been widely adopted by law enforcement agencies, resulting in wrongful arrests and prosecutions. AI systems tend to perpetuate and accentuate existing biases under the guise of mathematical neutrality. Such systems are all the more dangerous when used for detecting and preventing crime, as law enforcement agencies often have a history of discrimination and prosecution of vulnerable communities. Therefore, even with the proper transparency and governance practices in place, AI systems should never be used to make decisions impacting human lives and human rights in such a sensitive context.

DANGER: AI in surveillance

From tracking individuals to smart video surveillance, AI applications for security have proliferated these recent years. AI surveillance tools (as in, computer vision systems used in order to recognise humans, vehicles, objects and so on) are built into certain platforms for smart cities, remote sensing and smart policing (see the box on AI in policing). The use of these technologies to track and monitor citizens' movements and connections almost invariably results in huge privacy breaches and violation of human rights. For example, an investigation found that Clearview AI, a technology company, has illegally practiced mass surveillance of Canadians since it used billions of images of people collected, without these individuals' knowledge or consent, to train a facial recognition tool which was then marketed to law enforcement agencies around the world (Office of the Privacy Commissioner of Canada, 2021). AI systems have been shown to be exponential accelerants of preexisting surveillance practices, allowing for the questionable usage of data and unprecedented levels of control on citizens (Sahin, 2020). For example, the use of facial recognition technologies in the Alicem app, a national identity tool for online government services in France, raised an uproar from human rights organisations (Draetta and Fernandez, 2021). Furthermore, since technological evolution typically outpaces legislative changes and incoming regulations, it is becoming increasingly difficult to monitor such systems and fully understand the societal impacts they can have. AI surveillance tools can easily give way to harmful and oppressive practices such as the gathering of biometric information without consent, the manipulation of citizens' behaviour or the repression of ethnic minorities. Therefore, AI technologies should never be used in such contexts.

3.4.1.2. Supporting disaster response

HIGH IMPACT

LOCALLY RELEVANT

RISKS: Digital divides, geographic misalignment

Within disaster management, the response phase happens during and just after the phenomenon has happened. Depending on the situation, responses could involve evacuating threatened areas, firefighting, search and rescue efforts, shelter management or humanitarian assistance (Sun et al., 2020).

- AI has been proven useful for creating maps of areas affected by disaster events through remote sensing, which can help with situational awareness, to inform evacuation planning as well as the delivery of relief (Doshi et al., 2018; Bastani et al., 2018).
- By comparing maps and images pre-event and post-event, AI has been used to understand feature discrepancies and in turn, to assess damage of structures and infrastructures for prioritising response efforts (Voigt et al., 2007; Gupta et al., 2019).
- AI has been used to estimate the number of people affected after a disaster in order to provide efficient humanitarian assistance. Indeed, AI systems utilising location and density data for affected areas have been demonstrated to be a helpful alternative source of information, especially for precarious informal settlements.
- AI and satellite images combine to create new approaches for accurately spotting and differentiating structures in urban settings. This is especially useful in places that are informal, inaccessible, isolated, temporary or refugee settlements, or where buildings made of natural materials might blend into their surroundings (UNITAC, n. d.).
- For the same purposes, AI has been used for information retrieval on social media data. As an example, Twitter data can be used to gather information on populations affected by disasters as well as to geolocalise them.
- AI applications have also been used to identify the earliest warning signs of earthquakes, enabling emergency response teams to evacuate people faster.

RISK: Digital divides

During Hurricane Sandy in New York City (2012), Twitter served as a lifeline of information, helping affected citizens to spread information and facilitate safe and rescue operations. A study following the event, however, found out that most tweets spreading useful information about the disaster were geolocated in Manhattan, the richest region of the city, and the least affected by the catastrophe (Wang, Lam et al., 2019).

3.4.2. POPULATION RISKS

3.4.2.1. Assessing and mitigating health risks

HIGH IMPACT

Industrialisation and climate change are already having a concrete impact on the world's population exposure to health hazards. This is particularly evident when considering the ever-increasing number of heat waves in cities around the world, as well as the deterioration of air quality in highly industrialised countries. Such phenomena produce detrimental impacts on cities' populations, as prolonged extreme heat and pollution episodes can trigger chronic and acute respiratory diseases. AI can contribute to informing citizens and cities about health hazards in various ways:

- By utilising data collected through remote sensing, AI systems can provide insights on urban heat islands, water quality and air pollution at a highly granular geographical scale (Clinton and Gong, 2013; Ho et al., 2014).
- AI methods and demographic data can be used to assess which parts of the population are mostly impacted by climate change induced health hazards. Such information can help local healthcare authorities to drive outreach (Watts et al., 2017).

3.4.2.2. Monitoring and ensuring food security

HIGH IMPACT

Extreme weather phenomena caused by climate change, such as droughts, as well as geo-political events, such as wars, are already heavily impacting crop yields all around the planet. This poses a threat to food security, especially within communities depending on such resources. In this context, AI offers multiple monitoring and mitigation solutions:

- AI can be used to distil information on food shortages from mobile phones, credit cards and social media data. Such systems represent a valuable alternative to high-cost, slow manual surveys and can be used for real-time forecasting of near-term shortages (Decuyper et al., 2014; Kim et al., 2017).
- AI can also help with long-term localised crop yield predictions. These can be generated through aerial images or meteorological data (You et al., 2017; Wang et al., 2018).
- AI can also contribute to monitoring crop diseases by allowing their identification through computer vision techniques and informing agricultural inspectors of possible outbreaks (Chakraborty and Newton, 2011; Rosenzweig et al., 2014).

3.4.2.3. Managing epidemics

RISKS: Robustness, transparency, privacy

AI has been demonstrated to be a valuable tool for disease surveillance and outbreak forecasting. Some of these tools have important implications for equity. Indeed, AI-based tools can help healthcare professionals make diagnoses when specialised lab equipment is not accessible. For further discussion on this topic, see section 3.6.

3.5.

Water and waste management

3.5.1.

WATER

3.5.1.1.

Classifying consumption patterns and demand forecasting

Water utilities use long-range water demand forecast modelling to design their facilities and plan for future water needs. As water supply systems become stressed because of population growth, industrialisation and other socioeconomic factors, water utilities must optimise the operation and management of their existing water supply systems (Jain and Ormsbee, 2002). In addition, water utilities need to improve their predictions of peak water demands to avoid costly overdesign of facilities. One critical aspect for optimising water supply system operation and management is the accurate prediction of short-term water demands.

- AI can be used for forecasting water demand through the data collected by digital water meters (DWMs).

Irrigated agriculture is one of the key factors responsible for decreasing freshwater availability in recent years. Thus, the development of new tools which will help irrigation district managers in their daily decision-making process about the use of water and energy is essential.

- In this context, AI can also be used for the forecasting of daily irrigation water demand, including in data-limited settings (González Perea et al., 2019).

3.5.1.2.

Water quality prediction and wastewater management

HIGH IMPACT

LOCALLY RELEVANT

AI can support water management in response to sudden pollution events and seasonal changes and in modelling complex pollutants. For instance, algae recognition technologies can help in modelling algal occurrence patterns and understanding the presence of associated toxins. Water quality is not only related to biological or physical-chemical factors, but also to the continuity of the supply with adequate levels of pressure and flow. Many water utility companies are beginning to amass large volumes of data by means of remote sensing of flow, pressure and other variables.

- AI can be leveraged to detect the amount and composition of toxic contaminants, which can increase the efficiency of waste management systems (see section 3.5.2.2) (Alam et al., 2022).

Access to clean drinking water is a major challenge of the modern era and one of the UN's Sustainable Development Goals. Water pollution caused by rapid industrialisation and population growth has emerged as a significant environmental challenge in recent years. The treatment and reuse of wastewater through the aid of AI offer a unique opportunity to address both these challenges.

- AI can be used for the modelling and optimisation of the water treatment process, such as removing pollutants from water. In particular, AI can be used to predict and validate the adsorption performance of various adsorbents for the removal of dyes (Tanhaei et al., 2016), heavy metals (El Hanandeh et al., 2021), organic compounds, nutrients, pharmaceuticals, drugs and pesticides (Bouhedda et al., 2019; Gar Alalm and Nasr, 2018).

3.5.1.3.
Water level monitoring

Events at the polar opposites of the water cycle, such as floods of varying intensities on the one hand and droughts on the other, can have devastating effects on society. AI can offer substantial help in the monitoring of such events. (For ways in which AI has been used to mitigate the effects of extreme water-related events, see section 3.4.) Ideally, the data obtained by the remote observation of water-related phenomena should be used by decision-makers to assess potential interventions. These types of AI applications have also been used to inform farmers and in turn optimise irrigation at field scale.

- AI algorithms (specifically, computer vision) together with data from satellites have been used to identify trends in precipitation, evapotranspiration, snow and ice cover, melting, runoff and storage, including groundwater levels. In this context, the use of physical sensors is not suggested, as they could be easily affected by environmental changes (Chandra et al., 2020).

3.5.1.4.
Predictive maintenance

AI algorithms can provide spatial information on the amount and type of water losses.

- AI algorithms can be used to perform a continuous calibration of the network, including analysing the structure of the errors (difference between the measurements and the model predictions) at each control point, and extracting information from the error patterns. Different types of water losses can also be distinguished—for instance, pipe leaks versus unauthorised consumption.
- For district meter area monitoring there has been increasing interest in using sensor data for abnormality detection, such as the real-time detection of bursts.

3.5.2.
WASTE

SDGs: 12.4, 12.5

3.5.2.1.
Forecasting waste generation

Municipal solid waste (MSW) management is a major concern for local governments working to protect human health and the environment and to preserve natural resources. The design and operation of an effective MSW management system requires accurate estimation of future waste generation quantities, which are affected by numerous factors.

- AI has been used to forecast municipal solid waste generation. Traditional solid waste generation forecasts use data on population growth and average per-capita waste generation. Such historical data, however, are non-linear and highly variable; AI has been demonstrated to be a good tool to handle this uncertainty (Abbasi and El Hanadeh, 2016).

3.5.2.2.
Optimising waste collection, transportation and classification

Municipal waste collection is a costly and complex process. Trucks often visit waste bins that are only partially full, which is an inefficient use of resources. Moreover, the cost of waste collection and transportation accounts for 60% to 80% of the total waste management system costs; hence, optimising the route of vehicles for waste collection and transportation can save time, reduce the running distance, vehicle maintenance cost and fuel cost, and effectively arrange vehicles and allocate human resources (Abdallah et al., 2020). Many factors can affect waste accumulation, which makes it difficult to predict the fill levels of waste bins. In this context, AI can help in multiple ways:

- To improve the efficiency of waste collection, AI can be used to detect the fullness of a waste bin by real-time monitoring of waste levels within bins. Intelligent detection of waste bin levels can reduce the driving distance of trucks, reducing both cost and greenhouse gas emissions.
- AI has also been used to forecast the collection frequency for each location, reducing unnecessary visits to locations with empty bins.
- AI can also be used to analyse the effects of changes in waste composition and density on truck route optimisation.

Traditional waste classification mainly relies on manual selection, which is both inaccurate and inefficient. With the development of AI, many approaches have been proposed to improve the accuracy of recyclable waste identification through various techniques. Computer vision has been particularly helpful:

- AI has been applied on image data to recognise various types of cardboard and paper. Similarly, AI has been demonstrated to be efficient in recognising different types of plastics.

3.5.2.3. Optimising and controlling treatment and disposal

Quantifying useful by-products of municipal waste such as biogas and energy, as well as harmful by-products such as leachate and fugitive emissions, is essential for optimal waste management. In this context, AI can be used to predict the quantity and composition of different by-products generated from waste management processes within landfills during incineration and composting.

LANDFILL

Within landfills, AI can help with various types of prediction tasks, which in turn are needed for proper design and operation and can reduce environmental impacts.

- AI has been used to estimate landfill areas and monitor landfill gas, and it can predict landfill leachate generation using information on factors such as temperature and rainfall.
- AI has been used to predict biogas generation from bioreactor landfills. In this context, AI can help with the prediction and optimisation of the energy produced from solid waste fractions by using the physical and chemical composition of the waste.

INCINERATION

Determining the status of the municipal solid waste incineration process is a difficult task. This is due to varying waste composition, operational differences of incineration plants, and maintenance uncertainties of the incineration devices (Kalogirou, 2003). In this context, AI has been used for:

- Predicting and automating heating values during the incineration process.
- Monitoring and minimising the emission of pollutants in the air (Glielmo et al., 1999).

COMPOSTING

After composting, waste becomes a hygienic and odourless humus, realising the key aspects of harmlessness, waste reduction and recycling. AI can help model the complex processes that occur during composting, such as:

- Understanding waste maturity: AI has been used to understand the morphology, texture and colour characteristics of compost images and establish compost maturity.

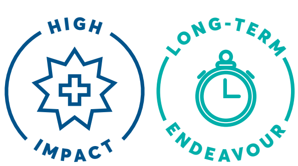
3.6.

Healthcare

Rapid urbanisation and environmental changes have had major health implications on city populations as global health challenges become more pronounced than ever. Furthermore, the double burden of infectious and non-communicable diseases can be exacerbated in the city environment (Galea and Vlahov, 2005). Many strategies have been implemented by local health authorities and their medical staff to improve population health. AI solutions for health have recently gained in popularity, as part of a wide array of digital health technologies (WHO and UN-Habitat, 2016).

3.6.1.

HEALTH PROMOTION AND DISEASE PREVENTION



SDGs: 3.4, 10, 5

RISKS: Privacy, data quality, historical bias, accountability uncertainty, reliability and robustness

AI can help support actions to enhance communities' health by improving care on two levels: efficient patient monitoring and management and relevant support for practitioners.

3.6.1.1.

Monitoring patients

Monitoring patients requires following up on patients' health condition and well-being regularly. This implies frequent exams and fluid communication between practitioners and patients. However, with the increasing pressure on the healthcare system and the inadequate coverage of territory, appropriate care cannot always be provided. In particular, marginalised areas of large cities and smaller cities with aging populations are those primarily affected by the phenomenon of medical deserts. AI represents a very promising tool to better engage with patients and enable remote monitoring of health conditions.

- Health monitoring systems (i.e., "remote health monitoring") may be used for continuous healthcare monitoring of users' body parameters: heart rate, body temperature, blood pressure, pulse rate, respiration rate, blood oxygen saturation, posture and physical activities.
- Self-reported health data can be collected with mobile devices and body sensors, such as wearable devices integrated into an IoHT (Internet of Health Things) or IoMT (Internet of Medical Things). AI techniques can be used to directly analyse this data and communicate results to the patient, who may adapt their behaviour (Chui et al., 2021; Nahavandi et al., 2022; Sujith et al., 2022).
- Real-time overviews of patient health status can be communicated to practitioners via an electronic health record, enabling them to act in a timely manner in the case of out-of-the-ordinary readings of vital signs (Chui et al., 2021; Santos et al., 2020).

3.6.1.2.

Supporting practitioners

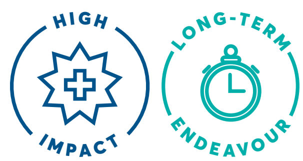
Medical practitioners face many challenges. A changing health landscape and increased pressure in medical fields can result in more complex responsibilities and lower professional well-being. Many practitioners are required to push past the limits of their skills in order to integrate new forms of information: multimodal, complex data such as images, genotypic data and numerical information. Furthermore, even the most experienced practitioners can be biased by their subjectivity. Both the quality of care and the wellbeing of the medical community can be improved through the support of various AI systems.

- Public health practitioners can be assisted by intelligent web-based applications and online smart devices which use AI methods to extract health and non-health data at different levels of granularity.
- Epidemiological information can be processed by AI techniques to provide evidence-based strategies of control of chronic diseases (Shaban-Nejad et al., 2018).
- Other areas of health intervention have been supported by smart systems. These areas include nutrition and diet, fitness and physical activity, sleep, sexual and reproductive health, mental health, health-related behaviours, environmental determinants of health and screening tools for pain (Cho et al., 2021).
- AI can be used for disease diagnosis and prediction: AI has successfully been integrated with other medical technologies such as medical imaging, including X-rays, CT scans and MRIs (Chui et al., 2021), for computer-aided diagnosis. These systems can identify areas of concern for further evaluation as well as provide information on the probability of diagnosis (Santosh et al., 2019; Zhou et Sordo, 2021). These techniques have been used effectively in relation to COVID-19, voice disorders, cardiovascular diseases, diabetes and especially Alzheimer's disease (Chui et al., 2021).

The nature of the application and the data integrated in the AI systems raises a series of issues pertaining to privacy and protection. Electronic health records represent extremely sensitive and confidential information on patients and their peers. Furthermore, the quality of the data may be questioned as well as its alignment and representativity of vulnerable populations and across genders. Accountability structures regarding AI algorithms are also central to the implementation of computer-aided systems in the health sector as poor technical performance may be wrongly attributed to health workers.

3.6.2.

HEALTH SYSTEM AND HEALTHCARE ORGANISATION



SDGs: 3, 11, 12, 16

RISKS: Skill shortage, financial burden, misalignment between AI and human values

High costs, workflow inefficiencies and administrative complexities are significant challenges the health sector faces which AI can help to improve. In particular, AI systems can support governments and agencies in the strengthening of the health system and the operation of relevant organisations (UN Committee on Economic Social and Cultural Rights (CESCR), 2000).

- AI can improve data entry and retrieval procedures through the automation of administrative and documentation tasks, including coding and billing, medication refill and reconciliations. Automated systems may also lead to a decrease in administrative complexities by identifying and eliminating fraud or waste, as well as assisting in the optimised allocation of human resources.
- AI can also be useful in predicting clinical key outcomes: prior insight on patient trajectories, mortality and readmission probability may improve the overall logistics and the management of hospital resources.

- AI could also be appropriate in the improvement of clinical decision-making by identifying optimal treatment choices. In some oncology applications, AI could be used to understand and address challenges such as poor clinical trial enrolment, disparities in oncological care, rising costs and non-uniform access to knowledge and multidisciplinary expertise (Lin et al., 2021; World Health Organization, 2021).
- AI has also been shown to be useful for complex decision-making, programme policy and planning. For example, AI has been used to predict, from administrative data, the length of stay of the health workforce in marginalised communities of South Africa. In Brazil, AI has proven to be advantageous in the allocation of resources across the country (World Health Organization, 2021).
- AI may also be used to measure the impact of health-related policies aimed at increasing workforce efficiency (Schwalbe and Wahl, 2020).

The success of AI integration in health administration and governance bodies will depend on the existing infrastructure and the financial resources a city has available for health investments. Without sufficient capacity to design, implement and use appropriate AI systems, a balance between their benefits and risks may never be found. In particular, implementing AI tools in an already pressurised sector may appear to increase employee workload because of the adaptation phase and the digital training required.

3.6.3. PUBLIC HEALTH SURVEILLANCE OF DISEASE OUTBREAKS

HIGH IMPACT

SDGs: 3, 11, 14

RISKS: Optimising the non-optimisable, data misalignment, geographic misalignment, mission creep

AI is instrumental in the prevention and management of public health threats, including the mitigation of, preparedness for and response to emergencies such as epidemics. The identification of early, accurate and reliable health indicators is crucial for the needs of health surveillance. Thanks to the generation of large amounts of health-related and population data, AI presents significant potential for the enhancement of health surveillance capabilities.

AI systems are able to source data to perform data analytics such as outbreak detection, early warning spatio-temporal analytics, risk estimation and analytics, and context-rich trend prediction.

Additionally, AI techniques can improve the quality of epidemic modelling, simulation and response assessments while considering complex interactions and constraints in the environment (Zeng et al., 2021). These prospective insights can inform authorities while comparing appropriate options for prevention and control strategies (Wong et al., 2019).

AI has already been used in the tracking of health behaviours during disease outbreaks and the spread of diseases.

AI can also be instrumental in the detection of outbreaks (Daughton and Paul, 2019; Gunasekeran et al., 2021; Schwalbe and Wahl, 2020; Xu et al., 2021).

However, although AI has evidently led to great advances in the field of health surveillance, reducing such serious issues as public health to technical solutions may lead to an overestimation of what AI can actually achieve on its own. The limitations of AI-powered models were highlighted, for example, during the COVID-19 pandemic as the predictive performances of many AI systems were questioned (Wynants et al., 2020). Furthermore, the quality of the predictions will depend on the quality of input data. The risk of poor demographic representation and geographic misalignment are prominent in a context of limited data and poor digital coverage. These risks are all the more important as these AI systems will be used to support policy decisions that have serious implications.

Case study: Kashiwanoha, South Korea

In Kashiwanoha, smart health strategies stretched the smart city agenda beyond technological innovations to address localised social issues. A mix of private and public actors, universities and citizens worked to implement AI experiments in monitoring and visualisation, educational initiatives and a variety of incentives for behavioural change. Thus, active pursuit of improved public health in Kashiwanoha has become a key part of the city's identity. These smart city strategies show how technology can create a link between the three sides of the sustainability triangle (i.e., environment, economy and society) (Trencher and Karvonen, 2017).

Case study: Khon Kaen, Thailand

In Khon Kaen, the ninth largest municipality in Thailand, the aging population is putting pressure on the municipality's public health capacity. To tackle this issue, the municipality developed the Khon Kaen Smart Health model, which incorporates three components: preventive healthcare service, smart ambulances and a smart ambulance operation centre, and health information exchange. Through the use of data gathered from IoT devices measuring various health factors such as heart rate and blood pressure, the project aims to give accurate predictions of vulnerable citizens' health conditions and personalised suggestions for better health outcomes. These personalised suggestions range from dietetic changes to increased sleep and increased physical activity. This made it possible to build an awareness of risk factors for non-communicable diseases among the general population, as well as provide advice and suggestions for proper behaviours. This initiative won the first prize of the IDC Smart City Asia Pacific Awards 2018 and the prize in the category of Public Health and Social Services (Godoy et al., 2021).

3.7. Urban planning

3.7.1. PLANNING AND MANAGEMENT

3.7.1.1. Population assessment

SDGs: 1, 10, 11, 16, 17

RISKS: Privacy and data protection, historical bias, data quality, data misalignment, presence of sensitive data

Understanding the social and demographic context when planning urban projects is crucial in order to produce sustainable designs that are not simply implemented in a top-down manner. Population assessment is therefore a requirement for the planning and management of cities. Machine learning can be used to measure, process and analyse a population's characteristics and behaviours for various objectives.

- Traditional machine learning methods are used to analyse official datasets from census or surveys carried out on the country or on the city level. These methods paint an aggregate picture of the socio-demographic characteristics of neighbourhoods to inform long-term policies and planning.
- The development of computer vision has provided the means of estimating population count and socio-economic conditions using remote sensing and GIS technology. This can be especially useful in contexts of very limited data and the absence of global data collection strategies (Xie et al., 2015).
- Population mapping is possible using new forms of data such as telecommunication data, credit card data and social media for real-time population estimation. The interactions between individuals and communities as well as their mobility behaviour can also be inferred. Information can be retrieved in order to identify meaningful neighbourhoods.
- Population mapping initiatives can be leveraged by AI techniques to evaluate the state of economic and spatial inequalities at the urban scale.
- AI methods can be adapted to include non-residents in the analysis of city behaviour for better planning. In particular, outside commuters or tourists can be included to provide richer information on the city.

However, because population data are very often of a sensitive nature, carrying out population assessment for the benefit of urban planning is not risk-free. Concerns over privacy and accuracy should be considered. Data collected through individual phone signals, credit card purchases or social media activities can reveal a person's location at all times, be explicit about their financial situation and reveal much about their identity. Very often the second-hand processing of this sort of data is not explicitly authorised by users. Many sensitive aspects of someone's personal life can be deduced from such data, especially when cross-referenced. This may lead to the marginalisation of or discrimination against population groups. Furthermore, the potential inaccuracy of those data and the resulting analyses should be accounted for, especially when leading to planning projects or policy-making. The digital divide in particular may often lead to a misalignment between the processed data and the affected population.

Case study: AI building tracker

UN-Habitat has piloted the use of AI to map the growth of informal settlements in eThikwini, South Africa, by detecting informal settlements on satellite imagery. Initially, there was no accurate data on the fast-changing informal settlements, but with the production of up-to-date maps through the use of AI, the local government is able to better plan and prioritise urban upgrading interventions. As a result, they are able to improve the delivery of basic urban services (UNITAC, n. d.).

3.7.1.2. Smart urban management

SDGs: 6, 7, 8, 9, 11

RISKS: Misalignment of AI and human values, societal harm, lack of mission transparency, shift in the labour market

The responsibilities of urban managers include the operation of many intertwined urban services. AI technology can be used to coordinate these different services in a more efficient way. Smart management refers to the use of AI by corporations to support local governments in planning and managing cities.

- Smart management centralises the management of various city-led services and infrastructure into a single virtual space, with some corporations providing additional data to help decision-making.
- Smart management tools serve to alleviate coordination and communication efforts between services, thus reducing the costs, time and risks associated with such complex tasks.
- The digital platforms used for smart management can present data visualisation and analysis to human agents, delivering key information on the status of the system in real time.
- Online "control rooms" can be implemented to oversee and combine larger sets of urban data.

However, this combination of expertise, strategies, services and data into what is defined as a "platform ecosystem" raises a series of concerns as to the risks of urban planning. Smart management tools have raised ethical issues regarding the influence of private companies on the objectives set by governments, while also questioning the reduced role of citizens in decision-making processes. Moreover, by replacing the role of clerks and other city employees, smart management heightens the risks associated with an increase in the automation of tasks previously performed by governmental officials.

3.7.1.3. Risk assessment and management

SDGs: 8, 9, 10, 11

RISKS: Misalignment of AI and human values, societal harm, lack of mission transparency, geographic misalignment

The population growth in cities and the increased risk of extreme natural events combine to produce very high negative impacts in urban areas. There is now a need to redesign the built environment to mitigate disaster (Caparros-Midwood and Dawson, 2015). The design of a resilient city requires proactively planning according to future risks and possible mitigation solutions. While AI tools are often focused on responding to adverse events (Sun et al., 2020), they can effectively support and complement preventive approaches as well.

In order to lay out a responsible and sustainable urban plan, city leaders can lean on AI techniques in the different phases of risk mitigation. City planners will first need to acquire a profound understanding of the natural context. This relates to knowing exactly what kind of risks the city may be exposed to. A knowledge of the existing urban context is then crucial for a design to be adapted to the effective needs. This implies linking the built environment and the population information to the specific risks identified. Finally, in light of the natural risks and the local environment, the spatial plan of the city can be optimised for more resilient urban forms. The design of infrastructure, regulations for land use and construction standards can be properly adapted for disaster mitigation. AI-driven systems can provide very interesting insight on these key phases.

- **Provide accurate maps of risk**: Local knowledge and historical data can be digitalised into a Geographic Information System (GIS) and coupled with remote sensing data for AI methods to better map out the potential geographic risks (Wang, 2018; Huang et al., 2021). For a comprehensive review of risk maps see Sun et al. (2020).
- **Create digital twins of buildings**: Related to both the ideas of digital twin of the city and building information management (BIM), various types of data can be used by AI to recreate models of buildings and ultimately building inventories (Wang, Lam et al., 2019). They can prove very useful when combined with population assessments in the next mitigation step.
- **Simulations using population assessment, digital twins and risk maps**: Relying on the previous models of the city (risk maps and digital twins), AI-based simulations can evaluate intervention scenarios in order to optimise the urban plan for specific objectives such as disaster mitigation.

The risks relating to these uses of AI will reflect those more generally encountered when integrating AI for public safety. More specifically in this context, the quality of the spatial information may not be sufficient to develop useful tools or may favour specific neighbourhoods for which data does exist. Furthermore, the rareness of extreme events makes for scarce historical data. To circumvent such limitations, learning from other contexts may be tempting but carries the risk of geographic misalignment.

3.7.2. NEIGHBOURHOODS

3.7.2.1. Real estate values

LOCALLY RELEVANT

SDGs: 10, 11

RISKS: Negative feedback loops, data misalignment, data quality, mission creep, historical bias

The rate of innovation in property technology through AI techniques such as virtual and augmented reality has been rapidly escalating in the last few years. Algorithms are being adapted as part of the city management tools and have been useful for estimating the market value of properties and forecasting future price trajectories of neighbourhoods. Very diverse information can be used to monitor the evolution of prices: property market data; property attributes such as size, quality and history; and area-level information such as mobility metrics, crime rates and public amenities. This allows managers to identify, intervene, control or predict the evolution of neighbourhoods and the advent of deterioration.

- Virtual immersive tours provide 3D views of real estate properties. They allow for zoom-in, audio commentary, and 360° viewing without physically being in the property area, which is useful for projects that are still under construction.
- Renters and landlords are finding AI monitoring and interacting dashboards to be a useful tool. These dashboards are able to monitor and offer everything from rent prices to inform landlords whether to raise or drop rates, chat boxes to allow them to communicate with tenants about repairs and maintenance, and automatically scheduled appointments based on the information provided by tenants (Cameron, 2018).

- AI algorithms are able to detect and identify potential fraud in real estate and insurance by allowing appropriate authorities to automatically read data from documents (i.e., statements, credit reports and proof of income) and efficiently allowing them to make a first round of decisions before they are passed on to underwriters.

However, the negative externalities of such systems on the fluctuating property market may not always be considered in their implementation. Therefore, acting upon the insight provided by AI may result in the deviation of outcome from expectations. Checkpoints must be included to keep track of the ongoing ethical issues to prevent the adverse effects of housing segregation, for example.

3.7.2.2.

Noise and comfort

NOISE

LOCALLY RELEVANT

SDGs: 9, 11, 13, 14, 15

RISKS: Data misalignment, data quality, concept drift

Noise is a leading source of discomfort for city residents and while urban noise seems unavoidable, techniques relying on AI exist to limit it. AI can ensure a minimum level of noise from transportation, construction, entertainment and human activity by providing soundscape insight for planners.

- Noise barrier optimisation in urban planning, for example, uses parameters such as absorption rate of barrier material, barrier height and other information on noise maps to simulate what measures need to be in place during the urban planning stage to reduce noise pollution (Trombetta et al. 2018).
- AI-powered measures have been used to address vehicle-based noise pollution and capture environmental noise anomalies. Data collection devices measure excessive noise levels on commercial or high-risk streets. Quantified notifications are sent to local authorities when a sound exceeds the decibel range determined on the street level (i.e., gunshot detection, number plate scanning) (Nokia, 2021).
- AI can be used in order to detect underlying patterns in the noise data that may represent early signs of natural threats. Researchers have developed a deep learning application that determines what ground noise is natural (i.e., what is human-made and what is not) and filters the data. Such AI systems can provide early warnings of earthquakes, for instance, as soon as the first tremors are detected. These seismometers—instruments that respond to ground noises—are placed in earthquake-prone areas (Yang et al. 2022).

Accuracy and discrimination should be a major concern when using AI in noise detection scenarios. Inaccurate or incomplete training data could result in either ineffective applications or negative bias against certain neighbourhoods. A deliberate effort should be made to make sure noise control and evaluation is as accurate as possible.

COMFORT

SDGs: 3, 11

RISKS: Geographic misalignment, outcome misinterpretation, negative feedback loops

Low-quality environments contribute to negative city settings in the form of a decreased sense of safety, vibrancy and liveliness. These aspects contribute to what can be defined as the comfort of an area. Urban comfort is the collective adaptation of a group of people in an area to certain microclimatic variables. Residents become acclimated to their outdoor urban space and define a range of parameters (i.e., thermal, visual, acoustic and air quality) at which comfort is achieved. AI techniques can encourage the inclusion of individual experience in the design of the city.

- Models are now able to use geospatial data to quantify and assess urban comfort levels via active monitoring devices that provide data to cloud servers. This type of monitoring system enables the generation of thematic maps. Indices of environmental comfort are then produced and can be used by residents (cyclists or pedestrians) when route planning or by local authorities for policymaking (Salamone et al. 2017).
- AI is able to predict and simulate the thermal comfort of urban plans before they are even executed. The thermal comfort of different urban landscapes can be determined so that planners can assess whether factors such as sunlight exposure, views and tree positioning affect comfort levels.

Urban comfort is part of the subjective experience of the city and, as such, is very much a relative concept: the way comfort is defined is context-dependent. Misalignment is a risk planners face if the system is not properly framed for the specific area and population. Furthermore, in order to limit the high energy consumption of these systems, there is a need to coordinate the management of all city sensors and urban services.

3.7.3. BUILDINGS, PUBLIC SPACES AND INFRASTRUCTURE

3.7.3.1. Smart buildings

SDGs: 11, 9, 12, 7, 3

RISKS: Digital divide, violation of privacy in data collection, concept drift

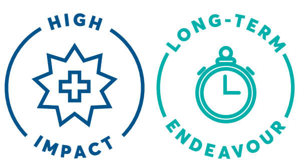
The comfort, security and energy consumption of homes and workplaces can be optimised when they are what we call "smart buildings." When smart devices powered by AI are included in the technical systems during the construction or renovation of the buildings, users are able to better interact with these systems and adapt them to their needs.

- Smart buildings, through embedded technological devices or smart objects inside or outside of buildings, can produce, process and transfer data and metadata about the property, its technical system and its users.
- Users can control and manage the technical systems of smart buildings remotely through commands or sensing.
- The building's systems can interact and self-calibrate using the IoT principle and optimisation algorithms.

Because smart building AI technology deals with domestic and work spaces, serious concerns exist regarding the transfer of information and data to third parties that own and design the technology or those that could access the data it produces. Audio and image recordings provide personal knowledge about the user's habits, characteristics and identity. Metadata and logs provide both precise and estimated information about occupants' routines, including their daily or long-term presence and absence. Additionally, individuals who have access to the data, such as co-occupants, can use such a system to monitor or watch other individuals. Furthermore, a key concern pertains to the centralisation of data and the concentration of information by dominant digital platform corporations that may influence urban planning decisions. The construction of smart buildings can also contribute to increasing inequalities between new and old built, digitalised or marginalised communities.

3.7.3.2.

Design



SDGs: 11, 12, 9, 7, 13

RISKS: Shift in labour market, lack of mission transparency, misalignment of AI and human values

Design has been integrating machine learning, deep learning and generative adversarial networks to test programmatic compositions and layout, develop building form and concepts, ease the production of technical drawings and 3D modelling, conduct historical surveys, optimise material usage and deepen our understanding of existing design heritage. The integration of AI systems in design, especially architectural design, is relatively recent and remains explorative (Chaillou, 2022). While it is often associated with a parametric approach in architecture, the use of AI systems encompasses a much wider array of applications.

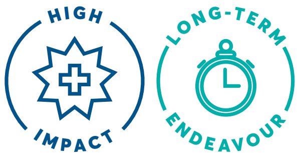
Many aspects of the design process could integrate AI systems. For example:

- Defining the implantation, bulk and shape of buildings
- Automating iterations of building program and floor plans layouts (As et al., 2018)
- Generating 3D models from 2D drawings or point cloud data
- Creating stylistic or ornamental pattern variations
- Defining the optimal distribution of green and blue infrastructure
- Evaluating the social response to building or open space design

While the use of AI would not replace the need for professional designers, it might greatly impact the labour market and work ethos of the design and construction industry. Designers and engineers might be required to validate their design or calculation using AI, creating factitious standards and expectations from clients. Such new requirements might put undue pressure on university programs in design or architecture to teach programming, with the risk of decreasing the perceived quality of schools that are unable to hire design professors with deep understanding of AI. The use of AI might offer a false guarantee for design projects that do not fulfil other quality standards of aesthetics, integration and innovation.

3.7.3.3.

Construction and structure assessment



SDGs: 9, 11, 12, 13, 17

RISKS: Skill shortage, shift in labour market, stacking of faulty AI, AI system expiration

Construction uses various AI systems, including machine learning, computer vision, knowledge-based systems and natural language processing (Abioye et al., 2021). Key applications include generating digital copies of existing or historical buildings and sites, streamlining project and budget management and automating construction work. The aim of AI applications is to deliver safer and more sustainable infrastructure faster and with reduced costs and risk. AI may also provide clues on past historical sites of cultural importance. Hence, the use of AI in construction relies on and helps produce many types of data such as 2D and 3D scans of buildings or building information management (BIM).

All phases of a building life cycle can benefit from AI—feasibility study, design, planning, construction, maintenance, renovation and demolition—for example by:

- Estimating construction costs and risks associated with types of ground
- Optimising the distribution of technical systems
- Predicting cost overrun from data about the design, the material and the workforce involved
- Automating the fabrication or the prefabrication of complex facade work
- Increasing the reliability of structural risk assessment
- Extrapolating on the lost patterns and artwork of historical constructions
- Predicting demolition waste (Akanbi et al. 2020)

Current use of AI in construction faces many limitations. The lack of reliable and complete data sets along with the complexity of combining material and human factors can lead to inaccurate or unreliable results. Construction sites and buildings are complex environments where human and material systems interact that are grounded in traditional knowledge and know-how. Changes in practices and job loss due to AI can lead to backlash but also lower the capacity of construction workers to maintain, train and transmit craftsmanship. Significant differences between buildings remains a barrier to the adoption of scalable standards in applying AI to construction.

3.8.

City governance

The different applications presented in section 3.7 highlight the opportunity that adapted algorithms can represent to help governments operate more efficient public services. Moreover, the administration of these urban services can be improved through AI as well. More generally, AI can be used to effectively support the management or governance of the city and better inform the decision-making process. This section presents AI applications that can be deployed by city managers on three different levels: enhancing local government capacity, engaging with the public and informing policymaking.

3.8.1.

ENHANCED GOVERNMENT EFFICIENCY

LONG-TERM ENDEAVOUR

SDGs: 12, 16

RISKS: Skill shortage, shift in labour market, outcome misinterpretation

A city's operation can gain in efficiency by integrating AI techniques for automating basic, time-consuming tasks. This can ultimately enable a better allocation of resources within the administration while decreasing the risk of human error (Berryhill et al., 2019).

- The integration of AI systems within the different public sectors has been used to encourage the standardisation of information systems and data-sharing processes in order to better coordinate operations. AI tools have also been used to design new operating models that foster interactions between services and build on their synergies (Anttiroiko et al., 2014).
- Software based on natural language processing or other automation processes has been used to generate reports, draft government documents, fill in forms and create visual communications (Lindgren, 2020).
- Better coordination of services through AI has enabled the data-driven optimisation of resource allocation by task automation and the identification of duplicated efforts (Zheng et al., 2018).

While the opportunities for reducing costs and human error are apparent, a combination of ethical and technical risks can arise when considering digital transformations on this scale. Notably, implementing these tools requires digital literacy on the part of civil servants and collaborators. Digital transformation within traditional administrations is a long process that goes well beyond the integration of AI tools. Not only can untrained employees feel marginalised and discarded, but being overwhelmed may lead them to misuse otherwise performant systems. Building the capacity of the government is necessary before undertaking internal AI transformation projects (see section 5.3).

3.8.2.

ENGAGEMENT WITH THE PUBLIC

The increased digital capacity of local governments integrating AI systems for service coordination or task automation can also positively impact the relationship between administrators and citizens. Efficient urban governance requires constant engagement with the public in order to ensure adequate consideration of their interests in the decision-making process. Interacting AI systems can be leveraged for meaningful engagement.

3.8.2.1.

Administrative processes

SDGs: 8, 10, 16, 17

RISKS: Violation of privacy, digital divide, geographic misalignment

Operational efficiency within government bodies can be applied to the different administrative processes that citizens must engage in when living in the city. In a digitalised setting, AI systems can support the redundant tasks of receiving or making payments and retrieving personal information or documents for application processes (Martinho-Truswell, 2018).

- AI-based systems have participated in creating secure entry points that connect people with all local government services and provide recommendations tailored to individuals' needs. Applications for government schemes can be automated based on a person's information shared across the different sectors.
- AI has been used to support the automation of administrative forms by using information from citizens' personal archives or historical data to generate typical responses (Mehr, 2017).
- AI is being used to support the application processes for social services by recommending resources that then need to be approved by human case workers (Barcelona Digital City, 2021).
- AI has been used around the world to provide a "digital persona" for citizens to facilitate online identification and access to services (Basu, 2017).

3.8.2.2.

Communication

SDGs: 10, 11, 17

RISKS: Data misalignment, violation of privacy, digital divide

The urban administrative procedures required of citizens are very much dependent on efficient communication between the local government and the public. AI provides tools that can accelerate information transmission and improve the quality of interactions.

- More personalised online communication has been enabled by AI systems such as virtual assistants and chatbots that can be used to direct users to the right services or to the right information (Safir, 2019).
- AI techniques have been used to assist citizens in their search for relevant documents, archives or material made available by the municipality.
- By linking personal and professional schedules, AI-powered tools are hosted by local governments to suggest administrative appointment schedules.
- Real-time translation enabled by AI has been used to help city officials communicate with the different communities that live in the city.

However, integrating AI in the interaction process between governments and citizens can entail the transmission of very sensitive, often confidential information through the system. Accountability issues therefore may arise if responsibilities and ownership of the tools are not clearly defined, in the advent of data breaches, for instance. Furthermore, transferring a majority of administrative procedures online can lead to poorer management of offline channels of communication. Negative consequences may appear where access to digital devices and digital literacy are unequally distributed across the population.

Building civic engagement and public trust

For a deeper dive into the use of AI and how it can affect civic engagement and public trust, we refer to a collection of essays curated by Brandusescu and Reia, written in parallel to the writing of this report in 2022.

Drawing from a variety of perspectives and across continents, the collection presents both civic and scholarly perspectives on engagement, partnership, law-making and new directions for urban governance. The essays, along with video recordings of the original symposium, are available in open access format (Brandusescu and Reia, 2022).

3.8.3. INFORMED POLICYMAKING

While local governments should lay out a clear strategy for the governance of AI (see section 5), AI solutions may be harnessed for the purpose of designing effective urban policies. In particular, the OECD advocates for the responsible integration of AI into the decision-making process to ensure that decisions align with the needs of the people and to anticipate their impact (Berryhill et al., 2019).

3.8.3.1. Identifying local needs

HIGH IMPACT

SDGs: 1, 2, 3, 4, 5, 10

RISKS: Outcome misinterpretation, data quality, geographic misalignment

Understanding the behaviour and interaction of agents in the city is the first step in design policies. Research-led AI tools can be used by local governments to improve their knowledge of the context and to identify the urban issues that need addressing or will need addressing in the future.

- The data collected from the different and coordinated urban services can be effectively used to carry out event correlation and causal analysis for better-informed decisions (Shibasaki et al., 2020).
- Data-driven risk assessments and risk measurement devices can be improved through AI techniques to better identify and prioritise issues.

3.8.3.2. Formulating and evaluating policy

HIGH IMPACT

SDGs: 10, 11, 16

RISKS: Outcome misinterpretation, lack of robustness and reliability, lack of explainability, high energy consumption, geographic misalignment

Adapted solutions need to be formulated and integrated into local policy. AI techniques can help shape people-centred policies and support governments throughout their implementation. Similar tools to those developed for urban planners in the design of a resilient city can be developed for a systematic prospective analysis of policies.

- Digital twins of the urban form and the various urban networks have been used by policymakers and urban planners to predict evolutions in the city and measure the impact of infrastructure policies on the different areas (see the “digital twins” definition box in section 3.2.8.2).
- AI can be used by cities to evaluate the relevance of potential policies. By harnessing historical population data and research-led artificial models of society, simulations that mimic the local social communities can be combined with digital twins of the built environment to carry out economic and social impact assessments and optimise local policies or social schemes.

SECTION 4

Risk Framework

4.1.

Risk overview

The risk framework provides an overview of the risks of AI, along with evaluation questions to assess them. The framework can be skimmed for a general understanding, but also provides the necessary detail to support more technical teams in starting a broad assessment of AI systems. The risks presented are not exhaustive; the framework focuses on raising awareness about common issues at the intersection of AI's technical and societal implications. The aim is to enable cities to draft their own strategies for a responsible use of AI for sustainable urban development.

The risk framework highlights the different risks throughout an AI system's life cycle, which is divided into five phases: framing, design, implementation, deployment and maintenance. Each risk is presented with a simple definition and real-world examples from different geographical locations. Graphics and links show the relationships between the risks.

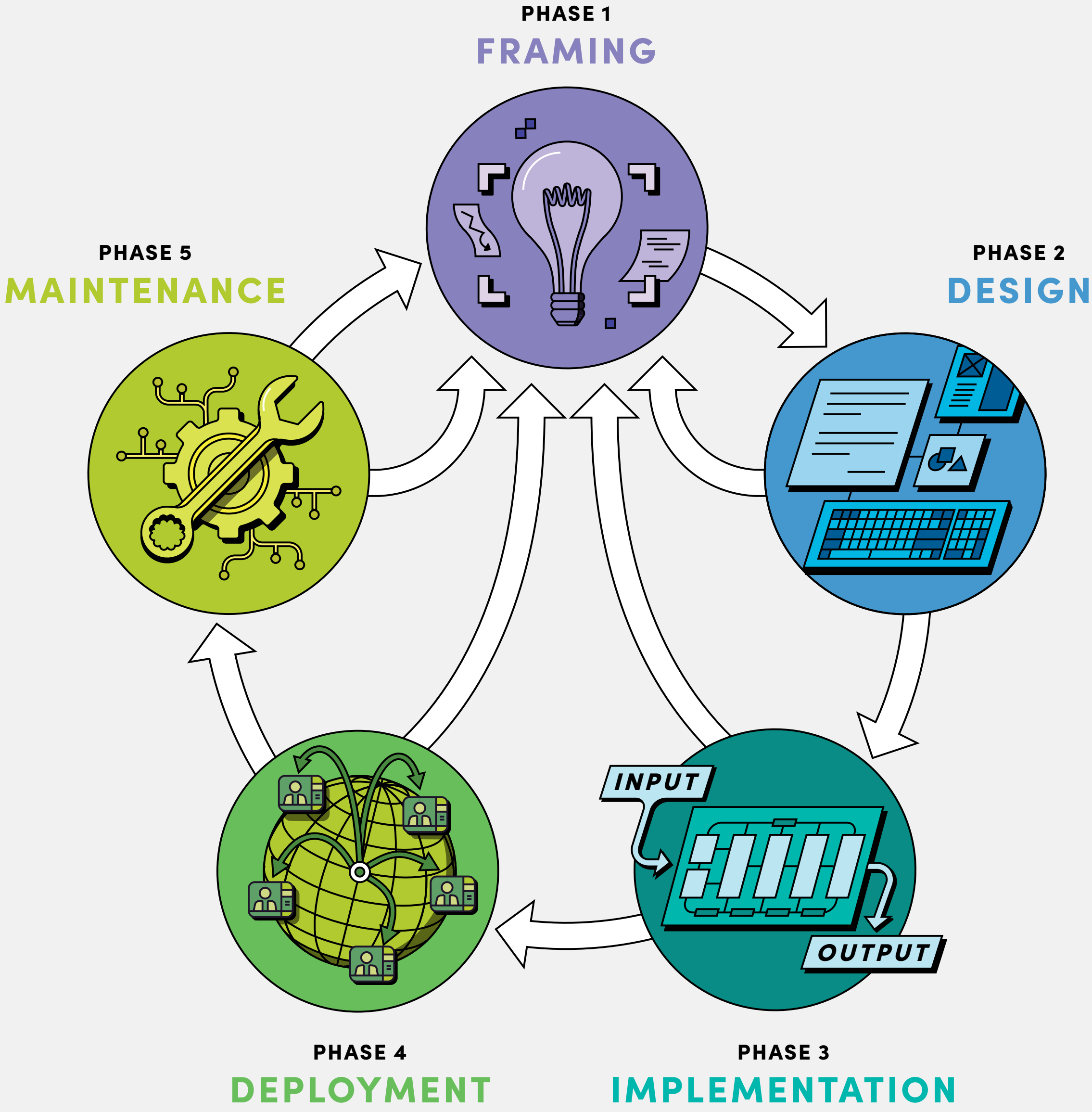
Each risk is accompanied by a series of guiding reflexive questions which function as an evaluation toolkit. By focusing on awareness-raising rather than on "techno-solutionism," the questions provide direction for locally appropriate mitigation strategies. The questions highlight places for possible intervention in order to mitigate risks while still being based in the specific context.

The assessment of potential risks in relation to AI systems must be done from a holistic perspective, encompassing both technical and societal considerations. Only when stakeholders clearly understand the structure and the limitations of an AI system will they be able to take full advantage of it and optimise the system's functioning in each particular context.

4.2.

The AI Life Cycle

The AI life cycle reflects the five major phases of an AI system and how it interacts with its environment. It is a tool that articulates the structure and processes of developing AI to help guide the reader's thinking. The phases are cyclical, and each one is intertwined with the others. For instance, the way an algorithm is designed and implemented defines its eventual deployment.



FRAMING PHASE

The **framing phase** is first. It focuses on problem definition and lays the foundation for later phases, which all point back to questions raised in the framing phase. This phase focuses on important reflections and risks arounds the context of an AI system's deployment.

DESIGN PHASE

The **design phase** focuses more on building the algorithm itself, before any coding. It builds on the parameters identified in the framing phase. The risks in this phase include questions about the team and the consequences to be considered, such as power and economic shifts.

IMPLEMENTATION PHASE

The **implementation phase** is the most technical one, and is structured around the AI pipeline. The issues of this phase are specifically about the algorithm itself and specific technical decisions.

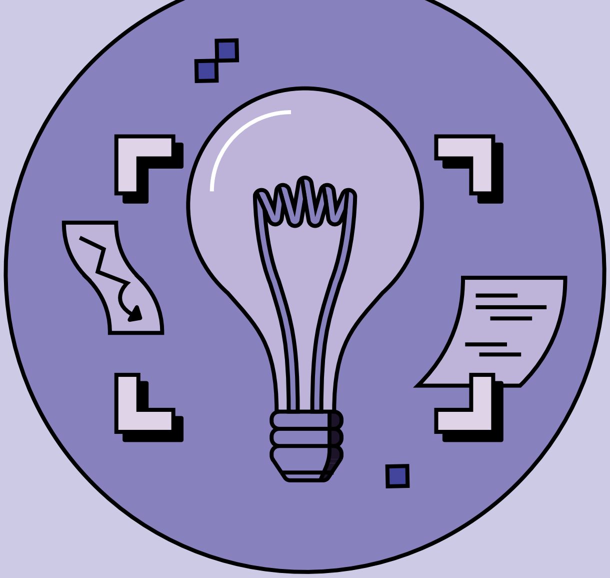
DEPLOYMENT PHASE

The **deployment phase** begins once an algorithm is fully developed. The risks from this phase arise when an algorithm is taken from a controlled and predictable laboratory setting to a real-world environment.

MAINTENANCE PHASE

The **maintenance phase** begins after an algorithm has been deployed and concludes when it is retired. This phase encompasses the long-term life of an AI system, including the considerations necessary to keep an algorithm functional and up to date.

In an urban context, multiple AI systems are combined and interact within an AI ecosystem. The complexity of the overarching system means early risk detection and intervention is recommended. For practical purposes, the different phases can also be connected to local project management processes to design concrete intervention points.



4.2.1.

PHASE 1
Framing

PHASE 1
FRAMING

PHASE 2
DESIGN

PHASE 3
IMPLEMENTATION

PHASE 4
DEPLOYMENT

PHASE 5
MAINTENANCE

The first decision for local authorities to make when considering the implementation of an AI system is whether or not to engage with AI at all. While the discourse surrounding AI promises increased efficiency, that may not always be the case. To determine suitability, one must first reflect on the challenge at hand, how it is currently being addressed, and how it may be handled better in that particular local context. Then, one should assess whether AI can in fact optimise a part or the entirety of the process (see section 2.2).

This stage is essential because when decision-makers engage in the framing process, they define the discourse and targets that will govern the creation and integration of an AI system. For an AI system to solve a problem, it must be given clearly defined, quantifiable instructions. As such, framing is important because the decisions made during this phase will contextualise and shape all the decisions that follow through later phases. Risk assessment in this phase depends on a detailed articulation of the problem for AI to solve, based on a series of social and economic factors specific to local context. Without this, any system which is developed can have serious flaws that impair its ability to improve the conditions defined within its objective.

4.2.1.1. Initial considerations

- Is AI the best tool to tackle this challenge? If so, why? What are the pros and cons?
- Have existing AI systems been applied in a similar context? What are the lessons learned?
- What is inadequate about current approaches to dealing with this challenge?
- How does the given mission relate to the challenge? How will it help solve the issue and to what extent?
- Do any system tasks involved require creative reasoning, such as complex human interaction? If so, an AI tool will not be likely to optimise the solution to the problem.
- How many system tasks involve gathering subtle cues or other information that can't be properly quantified? If there is a significant part of the mission that can't be quantified, AI won't optimise the solution to the challenge at hand.
- How does the AI system respect the best practices recommended by the responsible AI community (e.g., the Montreal Declaration for Responsible AI and the UNESCO Recommendations on AI Ethics)?

4.2.1.2. Framing risks

LACK OF MISSION TRANSPARENCY

The risk related to a lack of mission transparency arises when there is a lack of public communication regarding the objectives of an AI solution. In AI, transparency is considered a broad concept that stretches throughout different stages: it is related to algorithmic transparency, explainability, interpretability and trust (Larsson and Heintz, 2020). In the framing phase of an AI system, transparency mainly refers to the disclosure of information about the system (Li et al., 2021).

Municipalities must inform the public about why and how AI is able to optimise the solution to a public problem, how it is going to be applied, what the intended outcomes are, and what steps will be taken to achieve them (OECD and UN ESCWA, 2021). For instance, if police enforcement makes use of a face recognition technology for surveillance of public venues, it is important that it disclose how the technology is going to be applied, how it will collect data, how this data will be used, and whether and with whom it will be shared. In terms of the policy objectives, decision-makers should inform the public about what they are trying to achieve in the context of the mission or policy at hand and who is accountable for its application (Almeida et al., 2021).

Regularly engaging with the public throughout the life of the system encourages transparency and public acceptance. This implies providing access to information, conducting public consultations or other forms of citizen engagement from an early stage, and publishing the iterative process that led to the adoption of AI. By striving for social acceptance, city managers are ultimately working towards mitigating the risks of a fallout in the AI application, thus avoiding significant financial, reputational and social consequences.

Lack of mission transparency

RELATED RISKS

- Uncertain accountability
- Mission creep
- Regulatory breach
- Misalignment between AI and human values
- Lack of transparency and interpretability
- Outcome misinterpretation
- Violations of privacy in data collection
- Algorithmic aversion
- Unaudited algorithm purchase
- Concept drift
- Societal harm

QUESTIONS

- **What are the public policy challenges at hand? What sector and what stakeholders does it involve? Which communities are affected by it?**
- **What are the short-, medium- and long-term objectives of applying the AI system? How are those objectives aligned with the public policy challenges identified?**
- **Can the mission the AI will be given be explained in clear terms in a way that an algorithm will understand?**
- **Does the envisioned AI solution comply with transparency, interpretability and accountability best practices? (See the “lack of transparency and interpretability” portion of section 4.2.3.2 and the “lack of explainability” portion of section 4.2.3.3.)**
- **Where, how and for how long will the AI system be used? Is this information disclosed to the public?**
- **How can citizens easily access information about the AI-driven policy at hand? If such information is not yet disclosed to the public, how and when will it be?**

SKILLS SHORTAGE

In the context of framing what AI is used for and how it is used, the risk of skills shortage refers to the lack of human capacities. There are two common, significant limitations: the size of the workforce that is necessary to build and manage the AI system (human capacity) and the ability of this workforce to interact with and exercise oversight of the AI system (AI literacy). It is important to keep in mind the need for skilled professionals throughout the project, regardless of promises of automation. Technical support for maintenance at the local level creates an ongoing need for locally available skills.

Currently, there is a limited pool of AI talent due in part to economic or gender-related digital divides on the global and local scale, affecting primarily the Global South (WEF, 2020; Aguilar et al., 2020).

Beyond AI training, professionals must have cross-functional skills which allow for a proper optimisation of AI tools in the local context. Interdisciplinary and cross-functional competencies may also help avoid techno-solutionism by enabling a "human in the loop" (HITL) approach to algorithmic solutions. The HITL approach allows for an AI design that integrates human agency into certain critical decision-making steps of the system (see section 2.2 and recommendation #2 in section 5.2). There are instances for which the integration of human judgment improves performance, such as the balancing of fundamental rights. In fact, in high-risk applications, regulations may specifically require an HITL (Middleton et al., 2022; Mazzolin, 2020).

If these issues remain unaddressed, the AI infrastructure to be built and maintained risks replicating and perpetuating the inequalities represented by the skills shortage. Also, the lack of HITL risks leaving affected populations more vulnerable to a faulty AI system that is not optimised for the context in which it is deployed. Coupled with a lack of AI literacy, these effects can cascade into a generalised lack of trust in AI systems. Decision-makers should consider the extent of the human resources available to design, implement, deploy and oversee an AI system.

Skills shortage

RELATED RISKS

- Distress in the local labour market
- Inadequate infrastructure
- Financial burden
- Digital divide
- Lack of team diversity and inclusivity
- Geographic misalignment
- Negative feedback loops
- Unaudited algorithm purchase

QUESTIONS

- **What kinds of skilled professionals are needed for the mission at hand? Are these currently available in the region?**
- **Considering the human resources available, is there a skill shortage? If so, in what stage of the mission: data collection? Decision oversight? Model development?**
- **How will a lack of skilled professionals affect the design, deployment and maintenance of the AI solution envisioned?**
- **Can a potential skills shortage be overcome through partnerships that allow for a context-informed approach? If so, how?**
- **How important is it that the AI system envisioned be subjected to a human in the loop? What are the associated risks of fully automated decision-making and how significant are they?**
- **What is the worst potential outcome of the system? Will the system potentially lead to life-or-death or fairly complex situations to which a human should respond?**
- **Does the legislation demand a human in the loop? If yes, can the human oversight requirement be fulfilled?**

DISTRESS IN THE LOCAL LABOUR MARKET

The risk relating to distress in the local labour market arises when the adoption of an AI system creates structural changes in the labour market that yield negative effects on local populations. AI is not fully autonomous; its design, deployment and maintenance depend on human resources. While AI solutions may create new services and new forms of qualified or unqualified labour, they can also induce displacements in the labour market, as well as increasing the precarity or rendering obsolete certain trades.

In situations where workers are not replaced, their functions are often reduced to precarious unpaid or low-paid labour that requires them to execute microtasks that machines can't do efficiently. For instance, humans are needed to label images, as well as to translate and to transcribe texts, all of which are essential to support supervised learning techniques (Moreschi and al., 2020; Crawford, 2021, ch. 3). As another example, large-scale structural changes can be seen with how algorithmic platforms impact the mobility sector's dynamics by facilitating access to rides and food delivery services (Lee, Kusbit et al., 2015; Raval, 2019; Rosenblat and Stark, 2016; Kassens-Noor and Hintze, 2020).

Decision-makers should conduct a comprehensive analysis of how an AI system can impact present and future dynamics with regard to job opportunities, social assistance, economics and human development. Ultimately, AI systems only serve the population if they create a better environment for all. Although change may also imply positive spillover effects, municipalities must carefully weigh their ability to mitigate potential negative outcomes. It is important that these shifts be considered in the framing phase, as the effects of an AI system on the population may not be straightforward.

Distress in the local labour market

RELATED RISKS

- Skills shortage
- Regulatory breach
- Digital divide
- Misalignment between AI and human values
- Algorithmic aversion
- Societal harm

QUESTIONS

- **How does the AI system impact the job market? How does it impact different groups of workers and the quality of services provided?**
- **Are certain groups disproportionately more impacted by the AI system than others? If so, how?**
- **Does the AI system impact labour relations? Are these impacts determined by algorithmic decision-making? If so, how?**
- **What mitigation strategies can be applied to prevent distress in the labour market? Are they sufficient to balance potentially disruptive outcomes?**

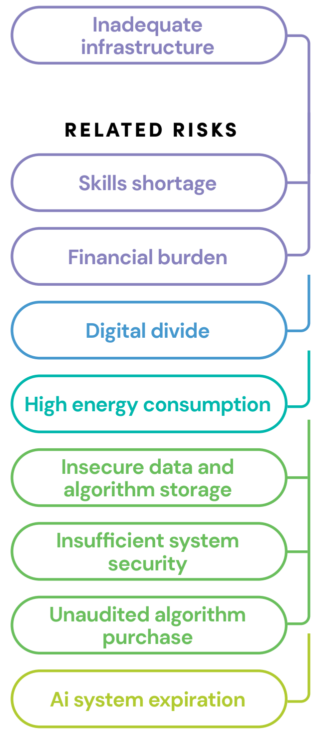
INADEQUATE INFRASTRUCTURE

The risk relating to inadequate infrastructure arises when the adoption of an AI system is not backed up by the level of technological infrastructure it requires for a safe and sustainable functioning. This risk stems from the idea that "it takes technology to make technology" (Bughin and Van Zeebroeck, 2018): AI systems depend on other layers of infrastructure, from ICT (information and communications technology) infrastructure to energy systems and hardware equipment. Designing and deploying AI requires the physical presence of broadband infrastructure with fast and reliable bandwidth. Furthermore, in order to do a great many calculations at high speeds, AI necessitates powerful computing resources, such as modern CPUs and GPUs (central processing units and graphic processing units) (Wu, Raghavendra et al., 2022). However, these processors can be expensive, with potentially volatile prices due to supply chain dynamics (J. P. Morgan Research, 2021; Rothrock, 2021).

While the computing power needed for AI can be outsourced through the usage of third parties' data centres, it is advisable for an AI to be developed and deployed in-house. For example, data processing centres located on a different continent could cause problems in terms of latency (delays in the amount of time taken by data to travel from a designated point to another). As another example, consider a situation where the AI deployer ignores a lack of data storage capacity to support an AI system's application in cities. This deficit can directly impact citizens' data protection (see the "insufficient privacy protection" portion of section 4.2.3.2) and result in security issues such as data theft and other forms of adversarial attacks (see the "insufficient system security" portion of section 4.2.4.1).

A "core" digital technology substructure must already be in place before the implementation of AI can realistically occur. This will most likely require previous investments in tangible hardware infrastructure, and more generally in the city's digital capacity. Municipalities must then consider what infrastructure they have to safely implement AI solutions, as well as whether the possible AI solution requires further investment in infrastructure.

Hence, infrastructure needs should be weighed during the framing phase, and investments should already be planned in the design of an AI strategy phase (see section 5). It is not advisable to implement AI systems on a promise of future investments and upgrades in the infrastructure.



QUESTIONS

- **Can the AI-related task be carried out with on-premises infrastructure?**
- **Is the available on-premises infrastructure safe and reliable? Does it meet current security standards and best practices?**
- **Do additional infrastructure needs have to be met before adopting the AI system?**
- **Who owns the infrastructure? Is it public or private? Are there any ownership constraints?**
- **If a public-private partnership is pursued, how is it going to work? Can this process be subjected to a risk assessment?**
- **Are data and data management protocols available for the envisioned AI system?**

FINANCIAL BURDEN

Risks relating to the financial burden resulting from an AI solution arise when decision-makers choose to deploy an algorithm they cannot afford to properly implement and later maintain. As suggested in previous sections, physical and human capital is needed throughout the AI life cycle, and these often come at very high costs (Davenport and Patil, 2012). Integrating an AI solution within an existing system and ensuring its sustainability through cyber-security will incur long-run costs for the public owners (Heemstra, 1992; Leung and Fan, 2002).

Should this constant investment in the system not happen, actors risk deploying dangerous systems into society. For example, if an AI system predicting catastrophic natural events such as landslides in the city did not benefit from the necessary financial resources to continuously maintain it, the lives of the people affected by the system's prediction would be at concrete risk because the prediction could be wrong; the accuracy of a deteriorating system will drastically decrease. Furthermore, the costs relating to managing the natural and social impact of failing systems may be significant.

Therefore, the investments will need to be protracted throughout the lifespan of the system. In essence, the financial capacity of a city should be balanced with the estimated costs of technological solutions before embarking on any procurement or design process. In particular, decision-makers should consider whether they will be able to sustain this investment in the long term and bear the unforeseen costs.

Financial burden

RELATED RISKS

- Skills shortage
- Inadequate infrastructure
- Digital divide
- High energy consumption
- Unaudited algorithm purchase
- Insufficient system security
- Ai system expiration

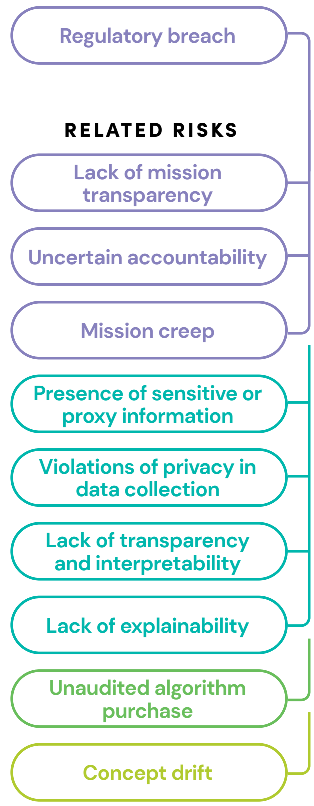
QUESTIONS

- **What is the estimated cost of the development and maintenance of the AI application?**
- **Will it be possible to minimise the financial investment needed without elevating other risks?**
- **Are there any resources available to address unforeseen costs? Is there an emergency fund dedicated to the AI system in place?**

REGULATORY BREACH

Risks related to regulatory breach arise when an AI system is incompatible with certain regulations of the jurisdiction in which it will be designed or deployed. Regulations are not limited to laws, but rather are a broad range of “instruments through which governments set requirements for enterprises and citizens,” including “laws, formal and informal orders, subordinate rules, administrative formalities and rules issued by non-governmental or self-regulatory bodies to whom governments have delegated regulatory powers” (OECD, 2018). In the context of developing, purchasing and deploying AI systems, a panoply of legal and administrative requirements might come into play. Ultimately, such breaches can jeopardise AI development and deployment, cut its life cycle short and have important financial impacts on local governments.

As such, the impacts of rules pertaining to each phase of the AI life cycle must inform decision-makers’ analysis of the advantages and drawbacks of adopting an AI solution. Before engaging with AI systems at any level—but especially at the framing phase—cities must undertake a thorough assessment of human rights and constitutional provisions, public-sector regulations, AI-focused regulations, privacy and data protection regulations, sector-specific regulations and different country-specific regulations in procurements of technologies designed in a different jurisdiction.



QUESTIONS

- **What regulations apply to the context of the mission and to the AI system’s application?**
- **How is the AI system being designed in accordance with the current legislation?**
- **How can regulatory risks be mitigated throughout the following phases?**
- **Does the AI system have a proven record of reliability and compliance?**
- **Is it possible to run an impact assessment and identify potential red flags for regulatory requirements and legal provisions?**

UNCERTAIN ACCOUNTABILITY

Risks related to accountability arise when there is a lack of transparency around the parties responsible for an AI system's deployment and maintenance. Within the scope of the framing phase, accountability refers to "the ability to determine whether a decision was made in accordance with procedural and substantive standards and to hold someone responsible if those standards are not met" (Doshi-Velez et al., 2017, p. 2).

It is important to note that accountability issues might arise whether the AI system is designed by public actors, co-designed across sectors or procured. For instance, governments may resort to public procurement to offer certain services to the public, or they may license an AI system application from a vendor. In addition to the public authority responsible for the AI design or purchase, various government agencies and even civil society organisations running social services on behalf of the city might be involved in some stage of the AI life cycle.

It is important to analyse the whole ecosystem around the technology at hand. Although the roles of the different stakeholders may be intertwined in a complex manner, it is crucial to be able to identify and map these in order to guarantee accountability. This is a necessary step for mitigating many other risks throughout the AI life cycle. Understanding the distribution of responsibility will be particularly useful to carry out the risk assessment for a system (Data Society, 2021). Governments and municipalities must take the lead in monitoring and enforcing measures for compliance. Ultimately, there should be a shared commitment by all actors to respect these measures.

Uncertain accountability

RELATED RISKS

- Lack of mission transparency
- Distress in the local labour market
- Regulatory breach
- Mission creep
- Presence of sensitive or proxy information
- Lack of transparency and interpretability
- Unaudited algorithm purchase

QUESTIONS

- Who are the stakeholders in the AI system's life cycle? What are their roles?
- Is accountability assigned in each phase of the AI life cycle? If so, how?
- Are all stakeholders able to justify their actions and the outcomes of their actions or omissions regarding their role in the AI pipeline?
- How are stakeholders throughout the pipeline able to redress any harms potentially caused by their actions or inactions?
- What public administration frameworks, mechanisms and resources (material and human) address accountability risks? Do these need to be adapted?
- Does the administration have a framework and best practices in place to respond to any event where it might be accountable for harms related to the design or deployment of the AI system?

MISSION CREEP

The risk of mission creep encapsulates the practice of deviating AI systems from their original purposes in a way that jeopardises their efficacy or underlying value system (Kornweitz, 2021). Mission creep is especially dangerous when a system initially meant for positive purposes is induced into serving other purposes that are likely to interfere with fundamental rights. The effects of such misuse can range from short-term unintended outcomes, such as discrimination and privacy violations, to long-term consequences related to the societal impacts of such practices. Overall, mission creep can be detrimental not only to the population, but also to the legitimacy of institutions and of technological development itself by increasing social distrust (Dwork and Minow, 2022).

An example of mission creep is the repurposing of a system initially invented to detect earthquake aftershocks for predictive policing. Evidence shows that the use of this predictive system to identify “hot spots” for criminal activity led police patrols to disproportionately target poorer areas, unveiling discriminatory outcomes (Mehrotra et al., 2021). Another example is the repurposing of a satellite imagery analysis system created for weather forecasting and prevention of landslides. When the system is used to map poor communities and conduct forced evictions of vulnerable families instead, the fundamental rights of those impacted are jeopardised (Greenfield, 2013).

As similar AI techniques can be used for vastly more applications than what they may be initially designed for, it is important to take into consideration its original application before assigning it other tasks. Furthermore, future repurposing options should be discussed when framing an original AI solution in order to limit the conditions in which irresponsible deviations may arise.

QUESTIONS

- To what purpose was the AI system at hand designed? In which context?
- What is the scope of the public policy challenge in which the AI system will be applied? (See the “lack of mission transparency” portion of section 4.2.1.2.)
- Is the AI system being used for the same purpose it was designed for? Are the contexts and purposes of the AI system compatible with the context and purposes of the intended application?
- Are there any differences between the scope of creation and the scope of application of the system?
- Can any such differences be mitigated? How so?
- Are the mitigation strategies sufficient to fill the gap between the AI system’s purpose and the policy context it will be applied to? If not, the risks of designing and deploying the envisioned AI system are likely to outweigh its benefits.

Mission creep

RELATED RISKS

- Lack of mission transparency
- Regulatory breach
- Uncertain accountability
- Manipulation and abuse through AI
- Misalignment between AI and human values
- Presence of sensitive or proxy information
- Lack of transparency and interpretability
- Negative feedback loops
- Unaudited algorithm purchase
- Algorithmic aversion
- Societal harm
- Concept drift

4.2.2.

PHASE 2
Design

The design phase encompasses the theoretical foundations of an AI system. During this phase, decisions about key aspects of the system are made, such as the data collection, choice of algorithm and outputs of the solution (Ugwudike, 2022).

Human beings are involved in all stages of the design process, from the problem formulation and outcome definition to the model construction. Their involvement is not neutral: assumptions are always made in the process of encoding human objectives into mathematical ones. These assumptions come in many different forms, often mirroring the social contexts in which the humans designing the algorithm find themselves situated. As such, the design of an AI system is heavily influenced by the designers' ideologies, values, theoretical assumptions and understanding of the task at hand (Ugwudike, 2022). The process of ideation is therefore a crucial phase in which human biases may be embedded in the system: at each step, choices that can lead to discriminatory outcomes are made (Leslie, 2019).

Biases against racial minorities, women and other groups that suffer historical discrimination can in part be explained by the designers' choices regarding which attributes to include or exclude in the algorithm. For example, the use of standardised test scores as attributes in enrollment and scholarship distribution algorithms can aggravate economic and racial disparities (Engler, 2021). In the United States, the use of such test scores as metrics for student success has been proven to perpetuate socio-economic inequalities, since students from less-privileged backgrounds score systematically lower (Smith and Reeves, 2020). Ultimately, the design choices may favour access to education for certain population groups over others.

More generally, decisions made during the design phase can lead to the creation of AI systems that replicate existing power structures, excluding minorities and widening socio-economic inequalities. Therefore, when designing AI systems, stakeholders should take into account not only the direct impact that their technology will have on the users, but also the indirect impact it could have on the surrounding socio-economic environment.

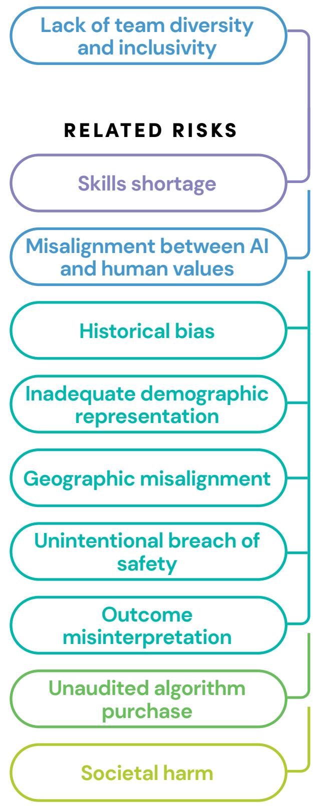
4.2.2.1. Design risks

LACK OF TEAM DIVERSITY AND INCLUSIVITY

The risk related to a lack of team diversity is that certain assumptions or decisions made in the design phase may not reflect the needs of all impacted communities. Lack of team diversity refers to a homogeneity of backgrounds—ethical, educational, cultural, religious and so on—among the professionals designing the AI system, whereas lack of inclusivity refers to a lack of decision-making power from various stakeholders throughout the AI life cycle. Together, the lack of team diversity and inclusivity can lead to negative impacts on society once an algorithm is deployed, most notably by perpetuating historically discriminatory practices.

This lack of diversity is regularly highlighted in the technology industry, where design teams tend to be small and essentially composed of men of middle to high social status (World Economic Forum, 2020). This recurrent scheme can result in unadapted technologies, because they are not explicitly designed for particular groups. For example, the supposedly comprehensive health tracker launched by Apple in 2014 did not even contain a period tracker (Criado-Perez, 2019).

To avoid such risks, local governments should ensure effort is being made to create a diverse team of designers. Moreover, to better understand the target population's needs, the different stakeholders that will eventually interact with or be impacted by the AI system should be included in the design phase. Very often, the people who are most affected by design decisions are the ones who have the least influence on the design process (Costanza-Chock, 2018). Furthermore, requesting the transparent disclosure of the team composition and of the design process is crucial to identify the underlying assumptions governing an AI system. Overall, decision-makers should ensure that the design of an AI system reflects and includes the same diversity as that of the population it will impact.



QUESTIONS

- **What diversity and inclusion practices are in place within the organisation designing the proposed AI system?**
- **How are the needs and perspectives of the target population addressed within the design phase?**
- **Has the team responsible for design identified the population subgroups which could be more at risk of discriminatory outcomes created by the AI system? What are the risks towards these subgroups?**
- **What are the principles guiding the design process of the new system? Has inclusion and diversity been integrated into the development of the system?**

MISALIGNMENT BETWEEN AI AND HUMAN VALUES

The risk of misalignment between AI and human objectives arises when values guiding the mission of an AI system are not reflected in the outcomes of the algorithm once it has been implemented. As mentioned earlier (see section 2.1.5), AI systems perform tasks with respect to concrete objectives. Therefore, in order for an AI system to engage with our world, a translation must occur between human and mathematical perspectives on the world (Korteling et al., 2021). Alignment means ensuring that AI systems capture the norms and values that guide human reasoning and motivate the use of AI. Misalignment can be broken up into two challenges: which norms to encode and how to encode them in the AI.

There are many values one may wish to encode into an algorithm's reasoning processes: privacy, safety, accountability and so on. A difficult concept to encode is fairness, which is used to distinguish beneficial from detrimental applications of AI. Multiple visions of what fairness is and how it should be measured exist and confront one another (Mehrabi et al., 2019). For example, demographic parity ensures that minority and majority groups are equally represented in the outcomes. Meanwhile, individual fairness ensures that two individuals with similar characteristics have similar outcomes. As such, although fairness can be translated into various mathematical concepts, it must be defined in a context-specific manner.

The technical difficulties associated with how to formally encode values, norms and human rights concerns into the AI (Gabriel, 2020) are not negligible. In 2018, a pedestrian was killed by an autonomous car while pushing a bicycle across the road. Investigations of the accident revealed that the AI system was not trained to respond to a person crossing the road at an unmarked location (Marr, 2022). Beyond the technical issues, there is a broader discussion surrounding which values should be encoded in an AI system in the first place. Many AI designs embody utilitarian principles, though this perspective is not universal and specifically contradicts many African cultural norms (Metz, 2021).

As our society becomes more entangled with various ecosystems of algorithms, the challenge of aligning human and algorithmic goals becomes increasingly complex, and it cannot be solved exclusively on a technical level. Decision-makers should therefore foster dialogues between computer scientists, ethicists, social scientists, jurists, policy experts and other domain experts to allow for an alignment between AI and human values.

QUESTIONS

- **What are the subgoals of the AI system's main goal? How are these related to each other?**
- **How are the goals of the AI system reflected in the design choices?**
- **Which goals are being prioritised when designing the AI system?**
- **Which values are being adopted when creating the AI system?**
- **Are the objectives guiding the AI system reflective of societal values?**
- **Are notions of fairness integrated into the AI objectives? Which ones?**

Misalignment between AI and human values

RELATED RISKS

- Lack of mission transparency
- Mission creep
- Lack of team diversity and inclusivity
- Digital divide
- Manipulation and abuse through AI
- Historical bias
- Inadequate demographic representation
- Geographic misalignment
- High energy consumption
- Lack of transparency and interpretability
- Lack of explainability
- Negative feedback loops
- Inadequate user feedback integration
- Societal harm
- Data drift
- Concept drift

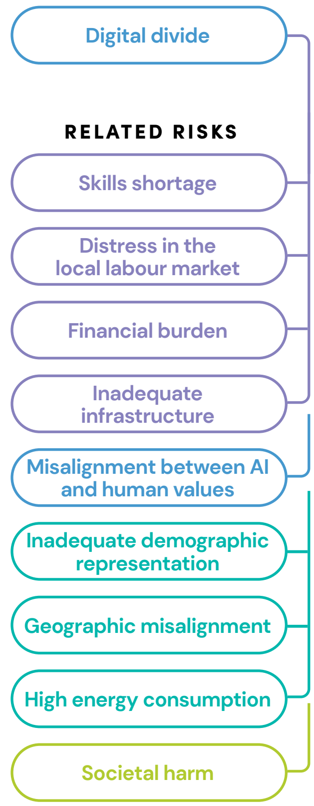
DIGITAL DIVIDE

Digital divides refer to substantial gaps in the accessibility of new technologies (Dijk, 2006). When they design an AI, relevant stakeholders should be aware of the risk that the envisioned AI system exacerbates existing inequalities. Digital divides can arise from a lack of access to the physical infrastructure that is needed for AI systems, a lack of digital skills, difficulties in accessing hardware, or users, in ability to obtain an economic return from the technology.

Beyond city dynamics, digital divides can be concretely noticed at a global level. The high sunk costs and large amounts of data required by AI innovation leads to the creation of monopolies: a small number of global frontier firms located in a few powerful countries serve the entire global economy (Korinek et al., 2021). The AI innovation race can therefore lead to a winner-takes-all dynamic, advancing countries that are early adopters and leaving behind most emerging economies due to a lack of adequate infrastructure, skilled labour and available data (Korinek et al., 2021). The existence of digital divides is the cause of multiple phenomena that perpetuate systemic discrimination and reinforce cycles of poverty. The lack of real-time access to information or the disparities in access to data and opportunities can prevent many developing countries from creating AI solutions adapted to the local conditions (University of Pretoria, 2018).

There are countless examples of how digital divides can materialise in individuals' everyday lives. For example, while a school might have computers for pupils to use, these computers might be obsolete and incompatible with the latest updates of relevant educational software, thereby impairing students' ability to benefit from the technology. In fact, schools in remote areas might not even have access to computers and broadband connection in the first place.

When designing an AI, stakeholders should be concerned with ways of closing the digital divide in order to ensure the newly created technology does not contribute to replicating existing inequalities (Eastin and LaRose, 2000). Moreover, municipalities should take specific actions to help those most affected by digital divides, for instance by investing in broadband infrastructure and digital literacy programs (Chakravorti, 2021).



QUESTIONS

- **Is specific hardware or software needed to use this AI system? Is it widely available? Who does or does not have access to this technology?**
- **How easy is this AI system to use, regardless of users' technological literacy?**
- **Could every municipality have access to and use this AI system, notwithstanding their geographical location?**
- **If the AI system's design directly impacts the population, how does it impact different geographical spaces in the city?**
- **Are specific socio-demographic groups disproportionately affected by the AI system?**
- **If this AI system aims to promote economic development, can everyone benefit equally, regardless of geographical location, socio-demographic group or disability status?**

MANIPULATION AND ABUSE THROUGH AI

The risks related to technological manipulation arise when design choices are intentionally meant to cause behavioural (or cognitive) changes in users' interactions with an AI system.

The massive quantities of personal data gathered for the purpose of prediction can enable AI systems to make predictions related to users' behaviour and digital consumption patterns. For instance, records of behaviour on social media platforms (e.g., Facebook likes) can be used to infer sensitive personal attributes such as religious views or sexual orientation, thus invading users' privacy (Kosinski et al., 2013). Many companies use predatory advertising and deceptive design tactics, also known as dark patterns, that enable them to influence users' choices and interactions with their AI system (Petropoulos, 2022). For example, targeted advertising isolates consumers and renders collective action against harmful or unethical products difficult. The same targeting practices could distract individuals from essential informational content such as healthcare announcements (Milano et al., 2021).

Given the innovative nature of the integration of AI systems into channels of communication, there remains a significant lack of public literacy on the ways in which these systems can impact users without their knowledge and consent. Furthermore, the accessibility of these systems exacerbates their reach and manipulative power. As such, decision-makers must carefully consider how the underlying design mechanism of an AI system might manipulate users into certain patterns of behaviour by exploiting vulnerabilities in their decision-making. Although these tactics can improve engagement, they can also be harmful to the societies they should be serving.

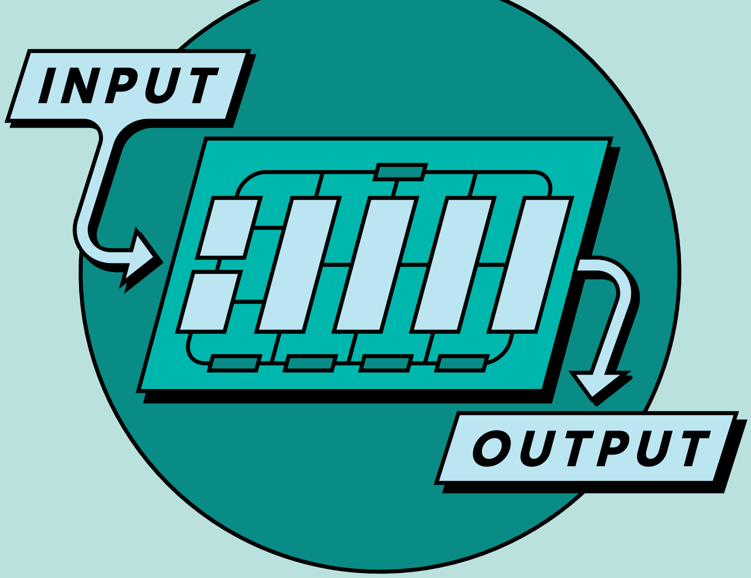
Manipulation and abuse through AI

RELATED RISKS

- Mission creep
- Misalignment between AI and human values
- Violations of privacy in data collection
- Lack of transparency and interpretability
- Negative feedback loops
- Algorithmic aversion
- Societal harm

QUESTIONS

- **Does the AI system aim to influence people into behaving in a certain way, or do so regardless of aim? How can this influence be examined?**
- **Does the technology facilitate the tracking, monitoring or influencing of people?**
- **Are the users of the AI system aware of the strategies in place to influence them?**
- **Have the impacted users provided consent before engaging with the system?**



4.2.3.

PHASE 3
Implementation

PHASE 1
FRAMING

PHASE 2
DESIGN

PHASE 3
IMPLEMENTATION

PHASE 4
DEPLOYMENT

PHASE 5
MAINTENANCE

The implementation phase corresponds to the process of building an algorithm. It can be divided into three building blocks: data input, the deductive algorithm itself and the algorithm's outcome. During data input, an algorithm is given training data which will inform its perception of the world. The risks within this process involve concerns related to the quality and the source of the data. During algorithm design, the algorithm itself is structured. The choices made in this process can define the complexity, interpretability, functionality and cost of the entire AI system. Finally, during outcome generation, the algorithm is given the input data and generates intended outcomes.

During the implementation phase of an AI system life cycle, the risks associated with the algorithmic process itself arise. Similar to the previous phases, one of the crucial elements of the implementation of an AI system is the alignment between the objectives among the actors engaging with AI systems. In the context of urban development, the values informing AI-driven strategies must coincide with human values (such as privacy, transparency, safety and fairness) in order to result in a trustworthy system (Li et al., 2021).

It cannot be overstated how crucial the implementation phase is to the successful engagement with algorithmic solutions. Although these risks are of a technical nature, they are directly linked to the earlier and later phases of the life cycle and can often be mitigated through appropriate framing and design choices. As the technical possibilities of AI are in constant progression, it is essential that stakeholders, in particular the ones that own and deploy AI solutions, be properly informed all throughout the stages. Namely, if local governments are aware of how technical issues translate further down the line into urban problems, they will be able to intelligently weigh the benefits and the drawbacks of using such systems before engaging in the tendering process. Furthermore, they may be able to implement the necessary measures to avoid adverse consequences. These risks are presented below with respect to each step in the general procedure for implementing an AI system.

4.2.3.1. Implementation risks: Data input

HISTORICAL BIAS

The risk of historical bias occurs when there is a limited understanding of the historical, socio-cultural and economic biases within datasets and the context in which they were made (Crawford, 2021). Data collection is more than a purely technical process, as it is shaped by human choices that are context-dependent and difficult to trace later (Rovatsos et al., 2019). Removing the data from its context of collection can therefore lead to harm, even when the dataset still reflects the world accurately (Suresh and Guttag, 2021).

Since AI systems require a large amount of data to learn, discarding historical data is not always feasible (Lattimore et al., 2020). Collecting more data to compensate does not mitigate the risks of unfair outcomes, since historical discrimination can still appear throughout the AI pipeline. Mitigating historical biases requires a retrospective understanding of structural discrimination (Suresh and Guttag, 2021); addressing historical biases requires more than technical solutions (Partnership on AI, 2021).

For example, the state of Oregon (US) has decided to retire an algorithm that screens for child neglect after it has been shown to disproportionately target Black families (Associated Press, 2022). This AI system used data without considering the contextual background where racial and income inequalities are closely linked. As a result, the AI system considered race as a factor for child neglect.

Data is not a neutral resource. Preventing the creation of AI systems that reproduce existing inequalities will only be possible by acknowledging existing discriminatory patterns and the socioeconomic, political and cultural norms that datasets represent.

Historical bias

RELATED RISKS

- Lack of team diversity and inclusivity
- Misalignment between AI and human values
- Inadequate demographic representation
- Presence of sensitive or proxy information
- Geographic misalignment
- Negative feedback loops
- Unaudited algorithm purchase
- Societal harm
- Data drift
- Ai system expiration

QUESTIONS

- **What is the time frame covered by the dataset? Does the dataset contain historical data? Data produced and collected recently?**
- **What are the potential social biases embedded in the data? What existing patterns of discrimination can be identified in the context surrounding the dataset?**
- **What information is available about the way the data was collected, labelled and pre-processed?**
- **Does the data collection process account for potential socio-historical biases? Can potential socio-historical biases be traced back and mitigated?**
- **How do the existing power dynamics and biases in the context of application exacerbate historical biases in the dataset?**

INADEQUATE DEMOGRAPHIC REPRESENTATION

The risk of inadequate demographic representation arises when datasets do not accurately represent diversity and groups are represented unequally. Population groups are defined by a set of characteristics, such as age or gender (see the "presence of sensitive or proxy information" portion of section 4.2.3.1). This risk can lead to deterioration of performance and perpetuation of discriminatory patterns. Excluding a population, although statistically "aligned" with the true population, has negative effects on both the performance and fairness of the AI system. These imbalances can be caused by undersampling or oversampling (not taking enough or taking too many data points from one group).

Poor group representation can have devastating consequences (Gebru et al., 2021). For instance, facial recognition systems perform poorly with minorities due to imbalances in training datasets (Buolamwini and Gebru, 2018). Many algorithms are unable to distinguish physical features among non-white ethnicities (Lao, 2020), and gender inequalities within biometric datasets have led to misdiagnosing diseases in female patients (Drozdowski et al., 2020). Poor group representation is often linked to inequalities in digital access, and data gaps can lead to spatial inequalities in urban services (Crawford, 2013). For instance, if a municipality seeks to improve road security based on feedback from a smartphone app, the representation will exclude neighbourhoods where many people don't use smartphones.

It is important to recognise biases throughout data collection as these will have cascading effects. Demographic groups must be represented with care while aiming for as much balance as possible. Especially when local authorities do not implement these algorithms themselves, it is important to share the local contextual knowledge they have in order to represent people's needs (see recommendation #2 in section 5.2).

Inadequate demographic representation

RELATED RISKS

- Regulatory breach
- Lack of team diversity and inclusivity
- Misalignment between AI and human values
- Historical bias
- Presence of sensitive or proxy information
- Geographic misalignment
- Negative feedback loops
- Unaudited algorithm purchase
- Societal harm
- Data drift

QUESTIONS

- **Which demographic groups are represented in the dataset used to train the algorithm?**
- **Does sufficient information exist about the dataset to understand its inclusions, exclusions and potential biases? Is there any proper documentation?**
- **Has the dataset been audited for demographic balance?**
- **Does the dataset include the entire population or is it a sample from a larger dataset?**

If the dataset is a sample from a larger dataset (Gebru et al., 2021):

- **What does the entire dataset consist of?**
- **Is the sample representative of the entire dataset? How was this representation validated?**
- **If the dataset is not representative of the entire dataset, how will you correct for the non-represented or underrepresented classes?**

GEOGRAPHIC MISALIGNMENT

The risk of geographic misalignment arises when data collected in a particular geographical context is used to train an AI in a different place. This risk, or portability trap, is common in data-scarce environments, especially in the Global South where pre-trained systems are often imported.

Geographic misalignments have serious potential to enact biases against local populations. For example, an AI system which is trained to assess loan eligibility in the context of a country with higher wages will show biases against a population with lower income in a different country (University of Pretoria, 2018). An autonomous vehicle calibrated for highway driving would not perform well in a chaotic urban environment where pedestrians cross unexpectedly (Gandhi and Trivedi, 2008).

It is important to make sure that geographic misalignment does not create conflict between the algorithmic system's objectives and the assumptions embedded in a dataset. City managers should be aware of the provenance of any system they choose to use in the local context and demand transparency with respect to the training process of the algorithms. Mitigating this risk will rely on the local expertise of the contractors and the understanding of the technological process.

Geographic misalignment

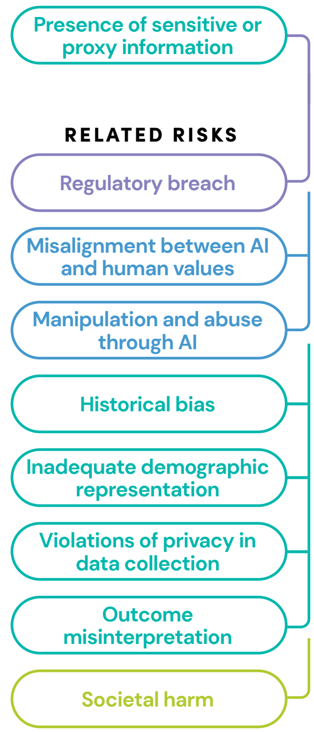
RELATED RISKS

- Lack of team diversity and inclusivity
- Misalignment between AI and human values
- Historical bias
- Inadequate demographic representation
- Unintentional breach of safety
- Lack of reliability and robustness
- Outcome misinterpretation
- Unaudited algorithm purchase
- Societal harm
- Data drift

QUESTIONS

- **Can the required dataset be gathered locally?**
- **If not, what are the differences between the context presented in the dataset and the one where the system will be deployed?**
- **Are the populations or classes represented in the dataset and those present in the environment properly aligned?**
- **How will classes missing from the training dataset but present in the local context be accounted for?**
- **What assumptions are made about the dataset for it to work in this particular geography?**

PRESENCE OF SENSITIVE OR PROXY INFORMATION



An important risk arises when sensitive information is used to train the AI as a basis for generating outcomes. A feature is considered sensitive if it about characteristics that are discriminated against, such as gender, age, ethnicity or sexual orientation (Amir Haeri and Zweig, 2020). This can create discriminatory patterns in AI outcomes.

For example, social welfare programs may choose to use a recommendation algorithm as support for a human caseworker. If the training dataset contains sensitive information about applicants, such as residency status, gender or marital status, the algorithm's outcome may base its suggestion on demographics rather than on information relevant for the welfare program.

Removing sensitive attributes is not necessarily enough (Prince and Schwarcz, 2020), as discrimination is often interconnected with many different aspects of someone's data. A dataset can contain proxy information which connects to sensitive information. Proxies can uncover hidden correlations, such as between social status and postal codes (Krieger et al., 2003).

Sensitive information will not necessarily pose risks. In medical datasets, it is often impossible to avoid the use of sensitive data. However, is it crucial to assess whether sensitive information is essential, whether developers and deployers have considered the downstream consequences of using this information, and whether there is a system in place to audit for discriminatory outcomes. It is important to engage in mitigation practices around sensitive applications, such as conducting a risk assessment on data usage and on which variables may become proxies. Data governance and strategy are needed to monitor data provenance and users.

QUESTIONS

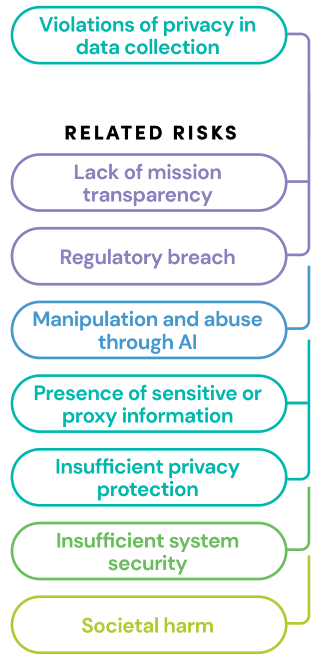
- **Which sensitive variables are present in the dataset? Sensitive variables may include information about an individual's gender, age, ethnicity or sexual orientation.**
- **What are the proxy variables present in the dataset?**
- **Are all the sensitive or proxy variables used in the dataset necessary for generating outcomes?**
- **Is the information in the dataset aligned with the intended use of the AI system?**
- **How are the proxy attributes measured to assess whether they are actually representative of the variables they are meant to represent?**

VIOLATIONS OF PRIVACY IN DATA COLLECTION

The risk related to privacy violation arises when data collection gathers information about individuals without their consent. Preserving privacy is critical; despite increasing regulatory data protection requirements, personal data such as purchase history information, credit scores, or even sexual or political preferences can easily be gathered without consent (Chui and al., 2018) (see the "presence of sensitive or proxy information" portion of section 4.2.3.1).

Public-sector AI systems often deploy in public spaces and result in collecting vast amounts of data from individuals. The more data collected, the higher the risks of unintended privacy violations, along with cascading impacts on fairness, regulatory compliance and security. Once the data is collected, data subjects often have no control over how it will be used. For instance, facial recognition tools adopted for street surveillance in Buenos Aires have collected sensitive personal data, including that of children, which was later integrated into a criminal profiling dataset (Hao, 2020). Collecting private data demands higher security protocols, as private information poses significant security risks to the civilians whose information has been gathered (see the "insufficient system security" and "insecure data and algorithm storage" portions of section 4.2.4.1).

These risks are much higher when data collection is done by an AI system. Algorithms are currently unable to detect bias, overrepresentation, imbalance, geographic misalignment and other data input risks, which raises the chance of potential unfair outcomes. Without human oversight, these risks are even less likely to be noticed.



QUESTIONS

- **How was the dataset collected? By which actors? With what consent?**
- **Is the collected data absolutely necessary? What is the minimum required?**
- **Was the data collection process tailored to the needs of the local population?**
- **Does the procedure of data collection comply with privacy guidelines?**
- **Is the data properly anonymised, containing only essential attributes? What privacy-preserving practices should be used to protect the collected data?**
- **Was the data collected by an algorithm? If yes, how was the data evaluated to ensure privacy protection? Who is responsible for monitoring whether the data collected is balanced in its demographic representation?**

4.2.3.2. Implementation risks: Algorithm design

LACK OF TRANSPARENCY AND INTERPRETABILITY

Risks relate to transparency and interpretability arise when decision-makers cannot understand the reasoning behind an algorithm's output, predictions or decisions due to its design. Algorithms produce outputs based on mathematically deduced reasoning, but as they are moved into more complex situations, it can be hard to decipher how deductions are made and based on what attributes (Dhinakaran, 2021).

The design choices made when selecting the architecture for an algorithm affect transparency and interpretability. There are many styles of algorithms which cannot be interpreted by design (Lipton, 2016). This lack of transparency can be very problematic if an algorithm begins to exhibit incorrect or unfair outcomes, because it is difficult to unravel what went wrong and extremely difficult to correct the problem. Algorithms base their reasoning on the intricacies of their dataset rather than on reality (see the "inadequate demographic representation" portion of section 4.2.3.1), so outputs can derive from irrelevant factors.

For example, an algorithm used to detect welfare fraud in the Netherlands was found to be discriminatory. It unfairly charged families thousands of euros, and the political cabinet resigned over the ensuing scandal. Investigations couldn't ascertain why the algorithm flagged a person for fraud if they owned multiple vehicles or garages. The rules of how these were codified were unclear, due to a lack of interpretability in the algorithm's design (Bekkum and Borgesius, 2021).

Therefore, decision-makers should carefully consider the risks of using architectures that do not allow for transparency or interpretability. While incredible innovations have been achieved on specific AI tasks, these successes have not included untangling AI's deductive patterns.

Lack of transparency and interpretability

RELATED RISKS

- Uncertain accountability
- Regulatory breach
- Manipulation and abuse through AI
- Outcome misinterpretation
- Stacking of faulty AI
- Lack of explainability
- Unaudited algorithm purchase
- Societal harm
- Concept drift

QUESTIONS

- **What are the inputs to the algorithm and how does it produce an outcome?**
- **How transparent is the architecture of the algorithm?**
- **What protocols are in place to enable independent audits of algorithms?**
- **Who is responsible for monitoring the output of the algorithm? How and when do they communicate with those designing it?**
- **How can various goals be managed and prioritised in a transparent manner?**
- **What is the contingency plan for monitoring the application's performance?**

LACK OF RELIABILITY AND ROBUSTNESS

Risks related to reliability and robustness arise when unexpected conflicts occur as an AI system is deployed. Reliability refers to the likelihood that a system will perform well in its real environment, while robustness refers to an algorithm's behaviour when something hinders its ability to function (Zissis, 2019). Although these risks arise during deployment, mitigation steps must occur during the implementation phase. Once the algorithm has been deployed, it is often too costly to make structural changes.

Robustness and reliability go hand in hand: In order for an AI system to be reliable, it must be robust to the variety of unforeseen factors inherent in real-world contexts. Unlike the controlled laboratory setting, in the real world, data may be incomplete, noisy or potentially adversarial. The ability of an AI system to react appropriately to such "abnormal" conditions and maintain operations during a crisis is broadly defined as "algorithmic robustness" (Xu and Mannor, 2012).

For example, an autonomous vehicle trained using a grid map of a specific city will inevitably encounter a variety of unexpected deviations when navigating a real environment. Robustness tests would make the algorithm account for unexpected deviations, such as road work closures. Reliability tests would demand that regardless of where the car began, it would successfully navigate to the destination.

There are serious consequences to a system which has been inadequately tested for robustness and reliability. The degree to which the robustness of an AI system should be assessed and prioritised can change depending on the autonomy and safety implications of the system (see the "unintentional breach of safety" portion of section 4.2.3.2).

Lack of reliability and robustness

RELATED RISKS

Inadequate infrastructure

Unintentional breach of safety

Stacking of faulty AI

Unaudited algorithm purchase

Ai system expiration

QUESTIONS

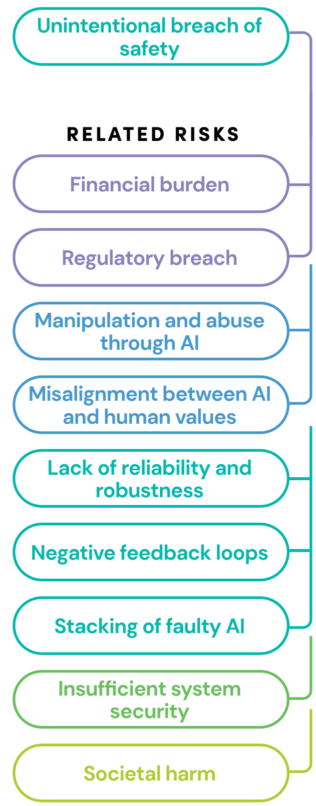
- **How does the system react to noise, i.e., when it is given many inputs with very slight differences?**
- **How does the algorithm react to anomalous data or environments?**
- **How has the system been tested for reliability and robustness?**
- **Have failures of the system been documented? What happens when the system fails or when failures are documented?**
- **What are the negative consequences that can arise from a system failure? Is there a backup plan for the occurrence of these situations?**

UNINTENTIONAL BREACH OF SAFETY

The risk of unintentional safety breaches arises especially when algorithms are deployed in safety-critical situations. AI safety includes the minimisation of the risk, uncertainty and potential harm incurred by unwanted outcomes (Varshney and Alemzadeh, 2017). Outside of safety issues related to adversarial attacks, the size and complexity of AI systems in an urban context make them vulnerable to human errors.

Errors can be catastrophic because an AI system has the potential to cause great unintentional harm if used improperly. Algorithms can cause harm even without being hacked, because maximising the wrong objective function (see the "misalignment between AI and human values" portion of section 4.2.2.1) can have unforeseen negative consequences when the system is deployed (Amodei et al., 2016). For example, a system which is used to map evacuation routes is highly susceptible to issues of safety if it is not properly monitored for the equal treatment of various neighbourhoods and their residents (Rohaidi, 2017). Accidents can also occur when the AI system lacks robustness. When deployed in a new environment, algorithms may exhibit poor performance because of data representation issues (see the "geographic misalignment" portion of section 4.2.3.1). Such AI systems can then commit harmful actions without even realising they are harmful, and as such, not raise any alarm (Amodei et al., 2016).

AI safety is highly connected to reliability and robustness because these elements make it possible to mitigate for risks related to new environments (see the "lack of reliability and robustness" portion of section 4.2.3.2). Having strong governance principles in place is an important step towards minimising the risks associated with the safety of algorithms (Falco et al., 2021) (see section 2.2).



QUESTIONS

- Is the AI system being deployed in a safety-critical environment? If yes, what is the procedure in place in case of accidents?
- Is human oversight integrated into the design of the algorithm?
- What negative consequences could arise from the desired application of the AI system? How are those negative consequences mitigated?

INSUFFICIENT PRIVACY PROTECTION

The risk of insufficient privacy protection arises when design choices made during the implementation stage of an algorithm leave it vulnerable to adversarial invasions of privacy. Beyond initial privacy considerations (see the "violations of privacy in data collection" portion of section 4.2.3.1), crucial privacy decisions also arise once the algorithm begins reasoning. Some technical choices, such as overparameterisation, can significantly increase the risks of privacy attacks against AI systems and their data points (Tan et al., 2022). This is because overparameterisation increases the chances of an algorithm absorbing detailed information from the dataset instead of inferring broad rules from data patterns.

The vulnerabilities caused by insufficient privacy-protecting design choices can be exploited in many different ways (see the "insufficient system security" portion of section 4.2.4.1). For instance, they may enable outcomes to be easily cross-referenced back to the memorised training source and, in doing so, reveal personal information about individuals. This is particularly concerning in situations where source data contains sensitive information, such as in the education or health sectors.

Privacy attacks evolve and can take multiple forms. Even with privacy-preserving design, due diligence is still required. Failure to account for the risks of privacy attacks against AI systems and their data can have spillover effects in other phases and, ultimately, on the lives and rights of the individuals concerned (Tan et al., 2022).

Insufficient privacy protection

RELATED RISKS

- Regulatory breach
- Manipulation and abuse through AI
- Misalignment between AI and human values
- Violations of privacy in data collection
- Insecure data and algorithm storage
- Societal harm

QUESTIONS

- **Do the inner workings of the algorithm save information regarding the initial training dataset?**
- **Is the algorithm's decision-making process secure? How do you know?**
- **Can the outcome of the AI system be used to retrieve information on the individuals involved in the training step?**
- **What are the mechanisms used to ensure that the input from users is kept private?**
- **What measures are being taken to secure identifying attributes, such as a user's location?**

HIGH ENERGY CONSUMPTION

An important concern with the implementation of AI systems is the energy consumption necessary, especially to power the training process of large systems. The choice of algorithm architecture also has significant impacts on energy consumption. Deep learning methods, which employ complex neural networks, have extensive carbon footprints associated with the high energy cost of training such a system (Gebru et al., 2021). It is difficult to predict exactly how much energy an algorithm will consume. Complex systems can end up demanding much more throughout their life cycle than was initially expected.

In cities, such increasingly high energy demands could not only incur increasing costs to the municipality and impact energy production, but also contribute negatively to the local environment in which they run. Potentially simpler algorithm designs may be more appropriate before moving to complex deep learning methods. It is also possible to integrate pre-trained algorithms, which allows system owners to avoid the prohibitive energy consumption costs of training a massive AI system from scratch (see the “unaudited algorithm purchase” portion of section 4.2.4.1).

Crucially, stakeholders should be aware of the direct impact that the complexity of a system will have on the amount of energy required to power it, from both a financial (see the “financial burden” portion of section 4.2.1.2) and an environmental standpoint.

High energy consumption

RELATED RISKS

- Inadequate infrastructure
- Financial burden
- Digital divide
- Misalignment between AI and human values
- Insecure data and algorithm storage
- Ai system expiration
- Societal harm

QUESTIONS

- What are the energy costs associated with training and deploying the AI system?
- What design choices can mitigate these energy costs?
- Do these energy costs outweigh the potential benefits offered by the algorithm itself?

4.2.3.3. Implementation risks: Outcome generation

LACK OF EXPLAINABILITY

Explainability refers to a process of deciphering an outcome, regardless of the design choices behind it. The risk of lack of explainability refers to the ability of human decision-makers to understand not only the outcome of an AI system, but also the variables, parameters and steps involved in the algorithmic decision process (Hussain et al., 2021). Transparency and interpretability are related to the design choices of the algorithm's architecture which disclose the AI system's reasoning.

The need for explainability arises when an algorithm is not designed in a transparent manner. As a consequence, the only way to explain the outcomes produced by the system is to trace back the rules guiding its decision-making from the interactions between the original training data and its generated outcomes (Thakker et al., 2020).

A lack of explainability can have a significant impact on the trustworthiness and social acceptance of these systems (Thakker et al., 2020) (see the "algorithmic aversion" portion of section 4.2.4.1). Stakeholders must be able to reason critically about the outcomes and functioning of an algorithm in non-technical terms. This is crucial for maintaining social trust (Beroche, 2021). Similarly, for regulatory purposes, it is often necessary for decision-makers to be able to justify outcomes. For example, in many countries in Latin America, public institutions are required to justify their decision-making process so that citizens have a right to challenge an outcome (Gómez Mont et al., 2020).

It is important to adopt explainability frameworks before deploying an AI system. Developers must be able to understand how the system works in order to identify and prevent problems from occurring, and policymakers must be able to understand potential biases or unethical behaviours that could arise (Thakker et al., 2020).

Lack of explainability

RELATED RISKS

- Lack of mission transparency
- Mission creep
- Misalignment between AI and human values
- Lack of transparency and interpretability
- Unintentional breach of safety
- Outcome misinterpretation
- Unaudited algorithm purchase
- Societal harm

QUESTIONS

- **Can the algorithm's inner workings and outcomes be explained?**
- **Are the outcomes of the algorithm used in safety-critical situations?**
- **Is there human oversight over the algorithm's decision-making process?**
- **Has there been an assessment of the alignment between the algorithm's performance and the desired outcomes?**

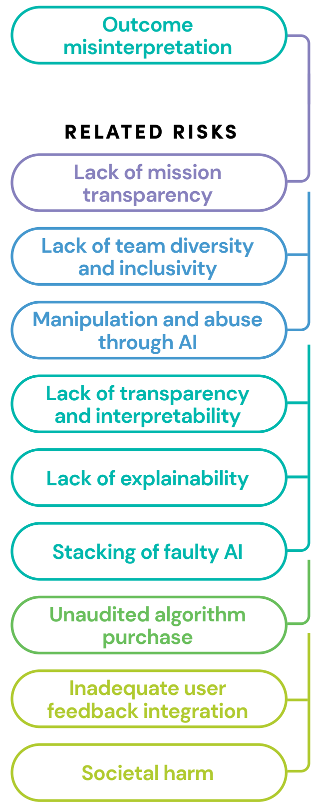
OUTCOME MISINTERPRETATION

The risk of outcome misinterpretation arises when human decision-makers must apply the outcome of an algorithm to their decision-making process. This risk is closely linked to explainability, but it occurs at the end of an algorithmic interaction, whereas explainability is related to understanding an algorithm's inner workings. This risk comes from the difficulty of interpreting how an algorithmic objective can be translated back to a human one. There are two aspects to outcome misinterpretation: lack of education on what an outcome represents and a blind trust in AI systems.

Lack of education becomes an issue when users who have no technical expertise are asked to use the predictions of those systems to make high-stakes decisions (Zytek et al., 2021). Depending on the formulation of a mathematical objective, it can be difficult to understand what the outcome actually means in that particular context. For example, social workers using an AI system for child welfare screening had difficulties using the outcomes produced by the algorithm because they didn't understand the output; while they used to make yes-or-no decisions, the algorithm provided a score from one to 20 (Zytek et al., 2021).

The tendency toward blind trust can aggravate issues of misinterpretation because that tendency can override a person's suspicions that the suggestions are not valid (Janssen et al., 2022). For example, tourists in Amsterdam were found biking within a highway tunnel because they had been directed to do so by a GPS system (Licheva, 2018).

It is important for decision-makers to mitigate outcome misinterpretation by promoting education, building capacities and exercising sceptical oversight of AI outcomes. In situations in which an AI system is given autonomous power without consistent human intervention, systemic risks could go unnoticed.



QUESTIONS

- **Are the conclusions reached by the AI system understandable for a non-technical person?**
- **Can the outcome of the AI system be directly used for decision-making or does it need to be processed first?**
- **Is there a direct relation between the outcome of the algorithm and the decision it is being used for?**
- **How are outcomes being monitored?**

STACKING OF FAULTY AI

The risk of stacking of faulty AI results from composition, or algorithmic stacking, which occurs when AI applications are combined. Although AI applications are often perceived as stand-alone products, in practice they are often integrated into a larger network of decision-making systems that work together. Algorithmic stacking can happen on different levels within a single AI system or when multiple AI systems are intertwined into an AI ecosystem.

The risks of stacking AI are twofold. First, if two algorithms that produce incompatible outcomes are integrated under the same system, they may not achieve optimal performance. Second, if many algorithms are combined, existing risks can be amplified or propagated. Privacy, fairness, explainability and other issues in a single algorithm could compromise the entire network (Dwork and Ilvento, 2018).

Ultimately, fixing all risks of a single AI system in isolation is not sufficient when multiple AI systems are used to build a solution.

Consider any autonomous vehicle: the self-navigation system is the byproduct of many individual algorithms, each pursuing their own particular outcomes and working together. One algorithm will be tasked with processing input from sensors, another integrating these sensor readings with the navigation control, yet another will monitor the speed. At the ecosystem level, consider the complex infrastructure behind an integrated mobility system for cities, including bike-sharing, buses, automated street lights, weather prediction and traffic conditions. Beyond each system's functioning, they are interconnected, with their functioning highly dependent on each system's performance.

Stacking of faulty AI

RELATED RISKS

- Mission creep
- Inadequate infrastructure
- Uncertain accountability
- High energy consumption
- Negative feedback loops
- Unintentional breach of safety
- Unaudited algorithm purchase
- Societal harm
- Ai system expiration

QUESTIONS

- **What are the downstream effects of integrating the new algorithm into an existing AI ecosystem?**
- **How compatible are the algorithms which have been used in the target application?**
- **How does the nature of each algorithm differ if the systems are used in tandem to produce outcomes?**
- **Have the algorithms within the AI system been audited separately or in tandem?**
- **Are procedures in place to address inconsistencies and incongruences between the algorithms' outputs?**

NEGATIVE FEEDBACK LOOPS

Negative feedback loops are a risk that is related to real-time algorithms which continue to gather data while deployed. These algorithms can augment their dataset with the responses they receive while interacting with their environment. One issue with this process is that its choices are participating in the building of its notion of reality (Liu, 2020). Intuitively, this is similar to an algorithmic version of confirmation bias (Nickerson, 1998).

An example of this negative feedback loop is illustrated in the problematic use of predictive policing (Hao, 2019). Imagine that police are dispatched to the locations indicated by an algorithm. A higher concentration of policing enables a higher rate of crimes to be reported. This suggests to the algorithm that more policing is required in that area. Consequently, we see a negative feedback loop where the algorithm's predictions begin to shape its reality rather than the other way around. This risk is exacerbated with the potential for biased outcomes (Jahnke, 2018) (see the "presence of sensitive or proxy information" and "violations of privacy in data collection" portions of section 4.2.3.1).

Another aggravating factor is when algorithms gather their own performance feedback. This is very common in recommendation systems, which either explicitly query users about their feelings related to the recommendations or infer them (Oard and Kim, 1998). Specifically in this context, there is a large potential for miscommunication between an algorithm and the user, because implicit feedback (i.e., the movement of a user's cursor or whether or not they clicked on a link) is "noisy," or not truly indicative of their underlying feelings. Over time, these misunderstandings can accumulate, instigating unintended changes in the algorithm.

Especially in contexts where an AI system is also integrated into an iterative data collection process, it is important to consider human oversight frameworks to monitor such AI systems that are likely to create negative feedback loops.

Negative feedback loops

RELATED RISKS

- Uncertain accountability
- Mission creep
- Misalignment between AI and human values
- Manipulation and abuse through AI
- Historical bias
- Inadequate demographic representation
- Presence of sensitive or proxy information
- Violations of privacy in data collection
- Lack of transparency and interpretability
- Inadequate user feedback integration
- Societal harm

QUESTIONS

- **How do the decisions made by the algorithm impact its environment?**
- **Is the algorithm using its own decisions as an input for the next round of predictions?**
- **Does the algorithm interact equally with different population groups?**
- **Is a process in place for monitoring the effects of the algorithm on its environment?**
- **Is a system in place for comparing the initial data distribution with the augmented one?**
- **Is the data collected by the AI system using explicit or implicit feedback?**

4.2.4.

PHASE 4
Deployment

The deployment phase of an AI system's life cycle includes all the risks associated with releasing an AI system into a real environment. Despite the mitigation efforts made during the implementation phase, some risks only come to light once a system has been deployed. Broadly, these are associated with security and societal acceptance.

Deploying an AI system at scale involves various levels of resources and financial planning. Appropriate computing and human capital are needed in order for solutions to be adopted widely and efficiently. Similarly, if an AI system is to be adopted widely, decision-makers need to make sure that the impacted population agrees with the AI in its form and scope. Lastly, an AI system needs to be secure from malicious attacks. In practice, malicious attacks represent one of the greatest threats to the wellbeing of a deployed system. Malicious actors can have a variety of motivations, from political reasons to an arbitrary desire to create mayhem.

For many municipalities, this may actually be the first phase in which they directly engage, if the algorithmic system has been purchased "out of the box" or procured. Regardless of whether an algorithm was developed or purchased, the risks associated with deployment must be appropriately addressed.

4.2.4.1. Deployment risks

UNAUDITED ALGORITHM PURCHASE

The risks related to unaudited algorithm purchase arises when decision-makers purchase a previously implemented algorithm and directly deploy it without considering the risks. There has been a noticeable rise in the purchase of "out-of-the-box" algorithmic solutions; most solutions are marketed towards medical (Davenport and Kalakota, 2019; Spatharou et al., 2020; Quinn et al., 2022), judicial (Rissland et al., 2003; UNESCO, 2020; Bench-Capon et al., 2012) and educational (Luckin et al., 2016) domains.

Many issues may arise from the purchase of a previously implemented algorithm, including those related to misalignment, transparency, safety and misuse (see the "mission creep" portion of section 4.2.1.2, the "misalignment between AI and human values" portion of section 4.2.2.1, the "geographic misalignment" portion of section 4.2.3.1, the "lack of transparency and interpretability" and "unintentional breach of safety" portions of section 4.2.3.2, and the "outcome misinterpretation" and "stacking of faulty AI" portions of section 4.2.3.3). Decision-makers purchasing AI systems must consider the degree of transparency required.

The owners of pre-made systems will always have less autonomy than those who develop from scratch; risk assessment during purchasing may be challenging as companies may refuse to reveal the details of proprietary algorithms. It is also possible that neither developers nor deployers will be able to make architectural modifications to the algorithm after purchase.

Consider the problems when local police departments in the United States purchased out-of-the-box facial recognition systems to identify crime suspects. Since they lacked the understanding required by the algorithm's design, they used forensic sketches rather than pixelated images as data inputs. Given this mismatch, as well as the critical context, the algorithm's deployment led to poor performance and discriminatory consequences on the ground (Garvie, 2019).

Unaudited algorithm purchase

RELATED RISKS

- Skills shortage
- Financial burden
- Mission creep
- Misalignment between AI and human values
- Historical bias
- Inadequate demographic representation
- Geographic misalignment
- Lack of transparency and interpretability
- High energy consumption
- Lack of explainability
- Data drift
- Concept drift
- Ai system expiration

QUESTIONS

- **Has a comprehensive risk assessment been conducted, addressing the risks outlined in this document?**
- **Are domain experts available locally to evaluate the efficacy of the system?**
- **What are the responsible AI practices of the AI system's designer? How has the algorithm been tested? How does the algorithm perform, what are its limitations, and how transparent is this information?**
- **How can the system be fine-tuned to best match the needs of the project?**
- **Is the code open source? Can it be modified? Who will do so?**
- **Does purchasing the system involve sharing local users' data with a private organisation?**

ALGORITHMIC AVERSION

The risk of algorithmic aversion arises when the society's response to an algorithmic solution impairs the solution's ability to perform optimally. Algorithmic aversion looks like the avoidance of engagement with, or even the boycott of, an AI system by the end users it was intended to cater to. Avoidance can cause societal unrest, undue harm and financial loss for municipalities. This risk can extend beyond a single algorithmic system.

Algorithmic aversion can arise when citizens are not sufficiently informed. For example, patients in hospitals show significant apprehension when informed that their diagnosis was made with the help of an AI system (Richardson et al., 2021). This avoidance can be exacerbated by unreasonable expectations of an algorithm's performance. Studies have found that people are quick to lose confidence in the ability of an algorithm once they have seen it make a single mistake (Dietvorst et al., 2015).

General lack of trust in society and governance can also contribute to the likelihood of rejection of a proposed algorithmic solution. Since AI systems rely on interactions with an environment to prove their use case, without the necessary engagement, an algorithmic solution cannot be assessed or improved. Trust in algorithms is a double-edged sword (see the "outcome misinterpretation" portion of section 4.2.3.3). It is important to cultivate a proper environment around an AI system so that it can be trusted, yet also properly scrutinised.

QUESTIONS

- **What are the positive and negative effects of the AI system on impacted communities? Will the system be integrated into citizens' daily lives?**
- **How will the positive and negative impacts of the AI system be communicated to the community?**
- **How has the impacted community reacted towards previous algorithmic systems?**
- **What level of public engagement will the proposed system require in order to satisfy its mission statement?**
- **Does the public have the knowledge to evaluate the new system with respect to their civic rights and needs?**

Algorithmic aversion

RELATED RISKS

- Lack of mission transparency
- Distress in the local labour market
- Uncertain accountability
- Inadequate infrastructure
- Mission creep
- Misalignment between AI and human values
- Manipulation and abuse through AI
- Violations of privacy in data collection
- Lack of explainability
- Outcome misinterpretation
- Unaudited algorithm purchase
- Insufficient system security
- Societal harm
- Ai system expiration

INSUFFICIENT SYSTEM SECURITY

The risk associated with insufficient system security arises when security vulnerabilities are exploited by third parties, through either malicious use or cyberattacks (PwC, 2018). Ultimately, a breach in security puts the personal information and lives of citizens at risk (Gómez Mont et al., 2020).

There are many places in the AI pipeline where vulnerabilities can be exploited. For instance, cyber-attackers can target training data through the use of data poisoning, where changes to the initial training set affect performance later (Newaz et al., 2020). By modifying the images of the traffic signs received by an automated vehicle, an attacker can make it behave unsafely and cause accidents (Ahmad et al., 2021). Similarly, hackers can attack the privacy of individuals through membership attacks, where the system reveals identifying information from users involved in training. A membership attack could be used, for instance, to expose patients' discharge from a specific hospital (Shokri et al., 2016).

On a system level, hackers can deploy a model inversion attack to reconstruct the deductive process of an algorithm (Zhang et al., 2021). The attackers can then create a fake version of the actual AI system (Krishna, 2020). Consider the consequences for the safety and wellbeing of citizens should such an attack be used on a grid used to monitor water usage across a city (International Telecommunication Union, 2020).

Insufficient system security

RELATED RISKS

- Regulatory breach
- Inadequate infrastructure
- Financial burden
- Misalignment between AI and human values
- Lack of reliability and robustness
- Lack of transparency and interpretability
- Algorithmic aversion
- Societal harm
- Ai system expiration

QUESTIONS

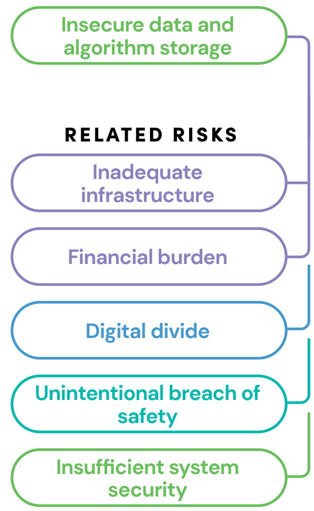
- **Has the AI system been tested for vulnerabilities? Have those vulnerabilities been documented?**
- **How secure is the system against malicious attacks?**
- **What is the contingency plan in case of an attack?**
- **What are the consequences of an attack? How severe are these? Who will be most impacted and how?**
- **How is the affected population going to be protected in case of a malicious attack?**
- **Should the algorithm be deployed considering the potential consequences?**

INSECURE DATA AND ALGORITHM STORAGE

The risk of insecure data and algorithm storage arises when the storage of any component within an AI system is outsourced to a centralised storage system. Distributed servers where massive datasets and complex algorithms can be stored are increasingly common because of the growing size of algorithms and their datasets (Sagiroglu and Sinanc, 2013).

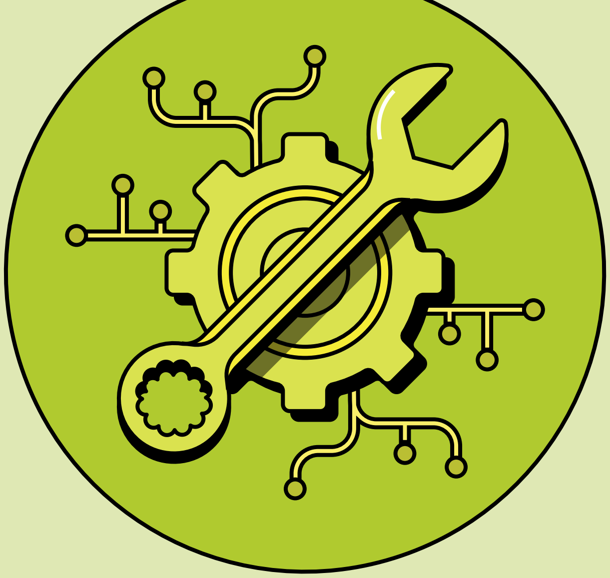
Cloud-centric architecture is often a contributing factor to the increasing costs and risks of running an algorithm (Khajeh-Hosseini et al., 2010; Lin and Chen, 2012) (see the "financial burden" portion of section 4.2.1.2). Storing private data on a cloud storage architecture which services many different clients can create data privacy and security risks. Thus, outsourcing storage often implies outsourcing security. Without the necessary protections, organisations storing their data on a server can be susceptible to ransomware attacks, where data is held hostage. For example, consider how ransomware attacks could halt food production if they were to target industrial farming grids that depend on complex AI systems (McCrimmon and Matishak, 2021).

Decision-makers who engage with cloud-centric architecture should assess the level of harm that could occur should these situations arise, and carefully evaluate whether server providers have adequate safety and reliability protocols.



QUESTIONS

- **How is the data stored? Can the dataset be stored locally?**
- **How large is the dataset? How much larger can it get over the AI system's life cycle?**
- **Who has access to the different datasets and the algorithm? Why? At what level of access?**
- **Whose data is being stored? What is the impact of a potential security leak? Who will be most negatively impacted and how?**
- **What security and privacy-protecting protocols are used to protect the data over its lifespan? What about to store the AI system's information?**
- **If applicable, what are the implications of relying on private infrastructure for computation and storage?**



4.2.5.

PHASE 5
Maintenance

The maintenance phase occurs after a system has been deployed and has been operating in its environment. The maintenance of AI systems involves monitoring how they interact with end users (such as citizens, residents and people in the city), the environment and the algorithm's objectives. The purpose is to maintain a connection between the values and mission objectives with the algorithm's actions over the long term.

Negative downstream effects are difficult to predict. While mitigation techniques are important, the extent and severity of a risk often surfaces only after the system has been released for some time. The risks of the maintenance phase relate to the consequences once an algorithm has been deployed into the real world for some time.

The maintenance phase describes a cyclical pattern that embodies the iterative process of algorithm design. Despite appearing to be last, this phase is interconnected with the rest of the AI life cycle. It is not unusual for an algorithm to go back through design, implementation and deployment phases due to risks that arise within the maintenance phase. As a result, structures that enable analysis and redress over time are necessary for truly successful engagement with an algorithmic system.

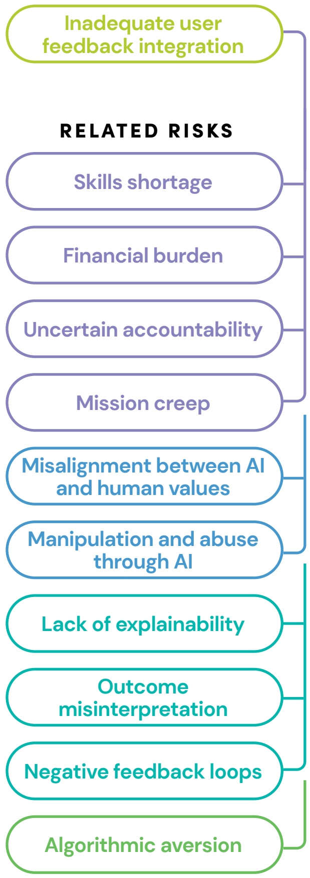
4.2.5.1. **Maintenance risks**

INADEQUATE USER FEEDBACK INTEGRATION

One risk that is unique to the maintenance phase revolves around a lack of action in response to user feedback. Developers can never assume that a deployed system will behave in the ways it was intended. Particularly in an urban setting where a large majority of algorithms are directly interacting with people in the city, integrating feedback is crucial. For example, consider how difficult it can be to assess a chatbot's performance in interacting with a broad range of dialects (Babyl, 2018). This risk arises when there is no structure to gather and integrate the feedback provided by those affected by the AI system.

There must be a format for gathering user feedback. Both providers and users of feedback must have a common understanding of the system's capabilities and limitations for feedback to be meaningful. Without this shared understanding, users would be unable to participate in genuine feedback, and providers would be unable to act on that feedback to make changes. As such, this risk is heavily tied to risks related to mission transparency, outcome misinterpretation, algorithmic aversion, accountability and lack of humans in the loop (Vathoopan et al., 2016) (see the "skills shortage" and "uncertain accountability" portions of section 4.2.1.2).

Finally, user feedback can be considered as one of the ways individuals can participate in the evaluation and redesign of an AI system. By engaging citizens with the outcomes of algorithmic solutions, policymakers can build trust while also improving the performance of an algorithm.



QUESTIONS

- **What type of user feedback does the AI system require? What are the negative consequences of not receiving feedback?**
- **How can users be engaged most effectively? What types of background knowledge are needed for effective participation?**
- **How do existing feedback processes affect people's trust in the AI system?**
- **Does the system explicitly prompt feedback? How effectively does the system track usage?**
- **What types of user feedback could be assimilated by existing learning algorithms?**
- **How is the algorithm calibrated for receiving and integrating user feedback?**
- **What kind of metrics and meanings are assigned to user feedback? What implicit and explicit information could be assimilated?**

SOCIETAL HARM

Risks relating to societal harm arise when a system presents widespread unintended negative consequences, particularly when risks from earlier phases remain unaddressed. Given the broad range of situations in which these interactions occur, these negative risks can materialise in emotional, behavioural and physical societal harm if left unaddressed.

For instance, in an example of misalignment between human and AI objectives, consider how predictive policing tools led to a disproportionate targeting of poor neighbourhoods (see the "mission creep" portion of section 4.2.1.2 and the "misalignment between AI and human values" portion of section 4.2.2.1). The use of that AI system may perpetuate feelings of danger and lack of trust among minority groups (see the "inadequate demographic representation" portion of section 4.2.3.1). Even worse, the increase in patrolling may impact targeted communities when altercations with police erupt into violent encounters.

The abuse of an AI system by users themselves is one unforeseen consequence. For example, social media platforms can be used to spread fake news through automated disinformation campaigns (Howard and Woolley, 2018; Vosoughi et al, 2018). Similarly, chatbots with female voices can be used to enable the practice of verbal abuse (Faggella, 2015). The downstream effects of such misuses of AI systems include emotional, behavioural and physical harm, the extent of which can be exacerbated by safety, reliability and robustness issues revealed only during maintenance. For instance, systems that are not properly audited can end up perpetuating harmful behaviours by replicating harmful patterns incorporated into their training dataset. As such, it is important for decision-makers to have frameworks in place for regularly assessing and auditing the performance of algorithms over the long term.

QUESTIONS

- **What are the mechanisms in place to evaluate, determine and detect societal harm?**
- **What mechanisms are in place to report misuse or concerns?**
- **How are the individuals interacting with the AI system being protected?**
- **How severe is the potential societal harm? Depending on this severity, should the system be in operation?**

Societal harm

RELATED RISKS

- Lack of mission transparency
- Distress in the local labour market
- Mission creep
- Lack of team diversity and inclusivity
- Misalignment between AI and human values
- Digital divide
- Manipulation and abuse through AI
- Historical bias
- Inadequate demographic representation
- Geographic misalignment
- Lack of transparency and interpretability
- Insufficient privacy protection
- Outcome misinterpretation
- Negative feedback loops
- Unintentional breach of safety
- Unaudited algorithm purchase
- Algorithmic aversion
- Insufficient system security
- Ai system expiration

DATA DRIFT

The risk of data drift occurs when the representation of the world in a dataset is no longer accurate. Data can become outdated or irrelevant due to large-scale societal changes brought on after the collection phase. Such changes can cause serious issues with the functionality of an AI system. One of the core assumptions of AI is that the dataset is used to guide future decision-making (see section 2.1.5). If the past data doesn't match the present situation, the algorithm will continue to rely on the data regardless and the system will lose predictive power (Saikia, 2021).

Data drift can happen in a single instance or slowly over time. For example, when an earthquake hit the city of Los Angeles, both the city topography and the future of construction changed drastically (Chandler, 2020). Any algorithm that had been trained on the dataset before this earthquake would be working with expired data. Similarly, the long-term degradation of sensors can affect an algorithm's capability to accurately perform environmental monitoring and forecasting (Ditzler et al., 2015).

It is important to consider the applicability of a dataset which repeats based on a context-specific cycle. Some datasets need to be updated more frequently than others. For example, a dataset used to train autonomous public transport may need to be updated at a faster pace than that of an algorithm used to monitor the effects of climate change on weather patterns.

Data drift

RELATED RISKS

- Financial burden
- Misalignment between AI and human values
- Historical bias
- Inadequate demographic representation
- Geographic misalignment
- Unaudited algorithm purchase
- Concept drift
- Ai system expiration

QUESTIONS

- What procedures are in place to account for changes in the AI system's context?
- Has there been a major change to the context or environment where the AI system is deployed?
- Is the AI system being used still relevant to the task at hand?
- How frequently should the training dataset be updated? What is the cost?
- How frequently should the AI system be tested for performance?
- What methodologies will be used to test for data drift?

CONCEPT DRIFT

The risk of concept drift arises when the properties of variables that an algorithm is trying to predict change over time (Lu et al., 2020). Different from the data drift risk, concept drift does not require changes to occur, only a re-interpretation of what they mean.

For example, consider an algorithm which is used to filter emails into a spam folder. As societal interpretations of spam definitions change over time, so do the algorithms that are used to make these predictions. Although the dataset itself may still be relevant, the concept of spam has changed. Similarly, if we consider an algorithm that is meant to identify harassment on a public forum, as the definitions of harassment evolve, so must the algorithms used to predict its occurrence. These changes will often require a dataset to be relabelled or replaced by a more reflective one. It can also require the re-training of the algorithm to replace outdated concepts.

Many AI systems are built on the assumption that the concepts presented in a dataset are stable over time. When this is not the case and a concept evolves, the reasoning of the algorithm becomes obsolete. For example, AI systems that were used to predict the quality of air using historical data were vastly thrown off by the lack of pollution during COVID-19 lockdowns (Mehmood et al., 2021).

Responding to concept drift during maintenance is crucial since it is only visible with time. As a result, it is important to adopt monitoring and control procedures for AI systems, especially in fields such as healthcare, governance and surveillance.

Concept drift

RELATED RISKS

- Lack of mission transparency
- Regulatory breach
- Uncertain accountability
- Mission creep
- Misalignment between AI and human values
- Geographic misalignment
- Lack of transparency and interpretability
- Unaudited algorithm purchase
- Data drift
- Ai system expiration

QUESTIONS

- **Are the theoretical assumptions on which the algorithm is based still applicable?**
- **What procedures are in place to test if the model still aligns with the objectives?**
- **How have the impacts of the changes happening in the surrounding environment been analysed and documented?**
- **Has periodic testing of the AI system been planned?**
- **How can resources be allocated should a re-training be required to ensure consistent performance?**

AI SYSTEM EXPIRATION

The risk of AI system expiration arises when an AI system which should be retired is maintained despite being problematic. Retirement means the ethical and efficient removal of an AI system. There are many reasons which can lead to the expiry of a system, including the risks in this framework if they remain unaddressed.

The risk of expiration arises at the final stages of an AI system's life cycle at different levels of urgency. The basis of expiration is the process of comparing the initial mission values with its current state of operating. Algorithms must be updated to align with the evolution of societal norms (see the "concept drift" portion of section 4.2.5.1).

In time, governmental changes or new regulations can require retiring an AI system (Madiega and Mildebrath, 2021) (see the "regulatory breach" portion of section 4.2.1.2). For example, recent bans on the use of facial recognition algorithms have forced the retirement of algorithmic systems in various countries (European Data Protection Board, 2021). Similarly, the decisions regarding the use of privatised datasets can cause architectures to become non-compliant. Systems which depend on banned datasets require replacement (see the "violations of privacy in data collection" portion of section 4.2.3.1 and the "insufficient privacy protection" portion of section 4.2.3.2).

It is the responsibility of system owners to have effective mechanisms in place for properly and ethically removing a system in its entirety at the time of expiration. This entails properly removing all infrastructure and destroying datasets according to the policies of use that should be stipulated.

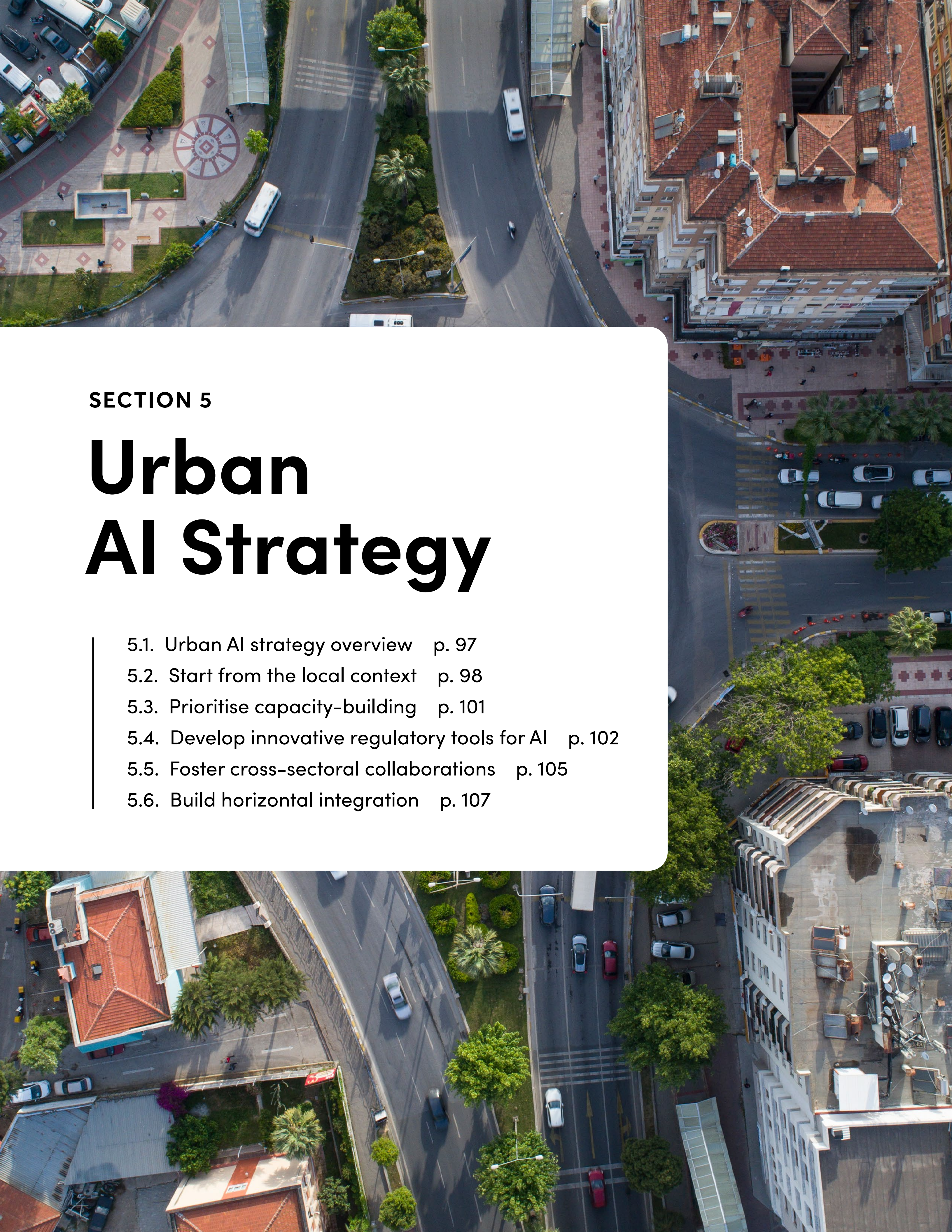
QUESTIONS

- Does the current AI system present concrete risks? If yes, are mitigation techniques in place to address them?
- How would a decision to retire a system be made? Has such a decision been made?
- What additional components are relevant to the functionality of the system? Will those be retired as well?
- What mechanism would be used to retract publicly shared trained AI systems or datasets?

Ai system expiration

RELATED RISKS

- Lack of mission transparency
- Financial burden
- Misalignment between AI and human values
- Historical bias
- Lack of reliability and robustness
- High energy consumption
- Stacking of faulty AI
- Unaudited algorithm purchase
- Insufficient system security
- Data drift
- Concept drift



SECTION 5

Urban AI Strategy

5.1.

Urban AI strategy overview

This urban AI strategy section is a how-to guide to help cities and local authorities develop AI systems that are in line with inclusivity and sustainable development goals. An urban AI strategy is the place to anchor the vision; it is a vehicle to articulate local, context-specific goals, as well as to plan actionable steps.

The section focuses on recommendations and concrete practical suggestions for local authorities on how to develop an AI strategy and governance framework. It includes considerations for building an enabling environment, fostering collaboration and building local capacity.

In addition, key tools that are specifically useful to support urban AI strategies, such as algorithm registers and algorithmic impact assessments, are highlighted in short case studies.

5.2. START FROM THE LOCAL CONTEXT

Each local context is unique, and AI systems must be developed starting from, and adapting to, the local context.

The successful deployment of AI systems is often determined by how the systems interact with their environment, as outlined in the Risk Framework (see section 4). It is therefore essential that a strategy be informed by the technical, political, geographical, social and economic context in which it will be deployed.

RECOMMENDATION #1: USE A PEOPLE-CENTRED DESIGN APPROACH TO AI SYSTEMS.

It is essential for citizens and communities to be involved in the development of an AI strategy. The first step is to engage the public. The active participation of a primary stakeholder—the public—will enrich the contextual knowledge and co-design of AI systems. Overall, engagement with the public through consultations, surveys, town halls and so on should lead to a more responsible and adapted AI strategy for the city.

To do this, it is important to clarify the strategic policy objective that a proposed AI system supports and to articulate how it will operationalise values in line with the public interest. Every AI system will embed values and assumptions, so it is important to consciously choose which values the system will support (see section 2.2). An effective AI strategy must develop a process to question the embedded values and assumptions in any AI system and its development.

The next step is to identify the affected communities targeted by AI systems, and then actually reach out to them and engage them through established community networks and processes. AI systems have a life cycle after deployment, and it is important to test the original assumptions to see how things actually work in practice and how communities are affected. This builds up evidence of the ways that deployed AI systems actually function, and can feed back into learning, monitoring and adaptation. And it will ultimately increase public trust in the use of AI-powered governments.

Case study: London

London has specifically identified collaborations between the public sector, the private sector and universities as a means of increasing the city's competitive edge. The City of London has engaged in a cross-sectoral collaborative city planning strategy, formalising the relationship between administration, industry and academia.

In the effort to manage data and AI research, both Connected Places Catapult and the Alan Turing Institute have partnered with the City of London. Their collaborations help start-ups and scale-ups based in London or operating there to develop their unique ideas. This partnership provides qualifying start-ups and scale-ups with new chances to collaborate with academics on data-driven urban challenges (Alan Turing Institute, 2018).

RECOMMENDATION #2: LEVERAGE LOCAL KNOWLEDGE AND EXPERTISE.

The successful implementation of AI requires meaningful interpretation and relies heavily on local domain expertise. Misleading conclusions can stem from failing to link both the input and the output of an AI system with local knowledge.

To prevent these risks, local knowledge can be included on two levels. The first is by incorporating local expertise and local types of knowledge in the process of shaping an AI strategy. This may include meaningful deliberative processes in the development of the AI strategy, for example. The second is by creating the conditions for which local knowledges may be systematically valued and included in future AI applications.

Different types of knowledge can contribute to shaping a public-interest-centred, context-based AI strategy. For example, tacit knowledge comes from the things we know from experience and practice, while contextual knowledge includes social and cultural norms and the way things are done locally (Ewijk and Baud, 2009; Buuren, 2009; Kitchin, 2016).

RECOMMENDATION #3: BUILD ON EXISTING INFRA-STRUCTURE AND DATASETS.

Existing resources can provide opportunities to draw on, as well as set limitations on what is possible. What AI systems can already be supported? What emerging initiatives can be further enabled? What financial and human resources are available? Does the city have stable internet access and reliable sources of power?

A key local resource is data. It is very beneficial to build an inventory that answers questions such as: What data sources are locally available? Which are accessible to the city, and which could become accessible, for example through data-sharing standards? What are the limitations or ethical issues in the data? (See the box entitled Case study: City of Los Angeles Mobility Data Specification in section 3.3.3.1.)

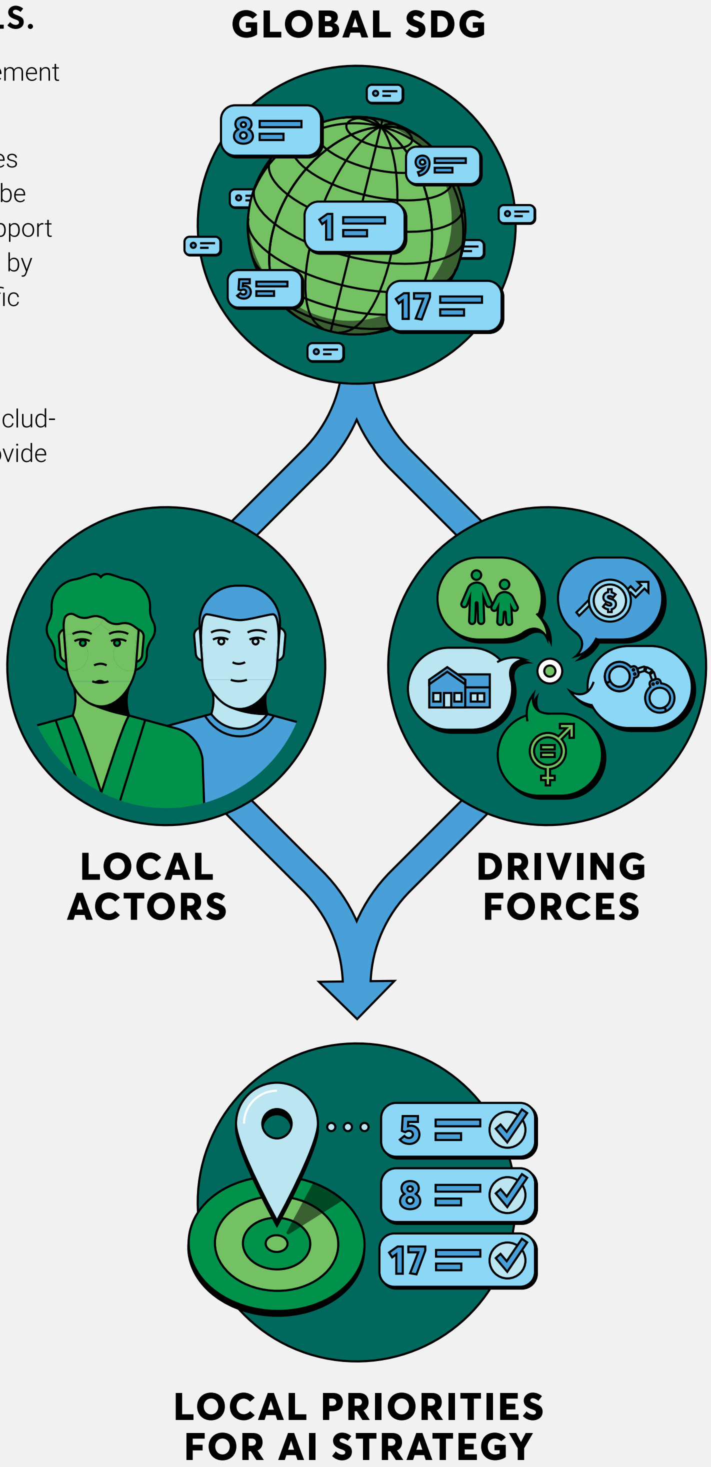
An important aspect to consider is the legacy systems of the city, which are previously existing technology infrastructures and databases. Cities often have to deal with maintaining dying technology infrastructures; upgrading or renewing urban services sometimes implies building on top of the existing systems. Aging software and qualified personnel turnover are common in all technology initiatives but the challenges they present are compounded in cities because of the population's strong dependence on urban services. A successful strategy must consider what systems already exist and how they may be adapted, upgraded or retired when the cost of maintenance is too high.

RECOMMENDATION #4: ALIGN YOUR AI STRATEGY WITH SDGS AND NATIONAL AND LOCAL GOALS.

How can the AI implementation support the achievement of the SDGs?

It is together with local actors and local driving forces that the global Sustainable Development Goals can be articulated at the local level. These priorities can support the choices of which issues to address, in particular by reflecting on the points in section 3 regarding specific SDGs that AI applications can target.

In addition to aligning with the SDGs, the AI strategy should also be guided by national and local goals, including specific measures to reduce unemployment, provide affordable housing or reduce carbon emissions.



5.3. PRIORITISE CAPACITY-BUILDING

Capacity-building is a significant element of any successful AI strategy. For an urban AI strategy, capacity-building is defined as the process of developing and strengthening the skills, instincts, abilities, processes and resources that organisations and communities need in order to plan, design and deploy AI applications.

For the public, capacity-building is about education regarding the opportunities, challenges and risks of AI. For local authorities, capacity-building includes investing in and providing opportunities for AI-related knowledge and skills development and attracting talent. It is essential that local governments create the conditions both to develop their own capacity and to build the capacity of its citizens.

RECOMMENDATION #5: EDUCATE THE PUBLIC.

As digital technologies and AI systems continue to transform everyday life, efforts to demystify and explain AI will play a major role in helping citizens understand AI systems and in building trust in an AI-empowered government. Increased awareness and knowledge of AI and AI application in the city will ultimately facilitate communication with the general public and with the private sector. It is important to recognise the diversity of audiences that need such education and to accommodate this diversity with a variety of educational strategies, taking into account, for example, different generations, levels of digital literacy, and so on.

RECOMMENDATION #6: INVEST IN INTERDISCIPLINARY SKILL DEVELOPMENT FOR EMPLOYEES.

Local authorities will need people with the skills to develop, design and deploy AI systems. While technical capacity is important, an entire ecosystem of interdisciplinary skills is also required for a thriving AI implementation. For example, AI regulation and law, AI ethics and AI business development are all key skills alongside computer programming.

As AI finds more and more useful applications in the city, the urban sector will increasingly use cross-functional teams; that is, teams that include a mix of skillsets. The ability to communicate across disciplines and to bridge perspectives so as to make the most of everybody's strengths and knowledge bases will be a key advantage for sustainable development.

In particular, cross-functional and interdisciplinary teams will be useful at all stages of project management for AI implementation, from procurement to maintenance. Each of the phases of the AI life cycle require reflection and evaluation, which is best fostered in these sorts of teams.

Local authorities will need to unblock the appropriate financial resources and create a conducive environment for these skills to thrive in the public sector. Training programs for employees across departments may also benefit the development of an AI culture within governing bodies.

RECOMMENDATION #7: INVEST IN BUILDING LOCAL TECHNICAL CAPACITY.

Cities must set aside training budget to upskill their staff on both the technical and business sides. All city staff members must have the necessary education about what AI is and how it changes current practice. A basic understanding and education about AI is required for AI implementation, particularly for procurement functions. When integrating an AI system in an urban sector, ensure that staff are trained and educated about the AI system that they are going to use. Make sure that the output of the system is clearly decipherable and applicable to their task. This requires technical and digital literacy for positions that may not appear to be technical at first glance.

RECOMMENDATION #8: DEVELOP DATA INFRASTRUCTURE AND STEWARDSHIP.

AI solutions require proper infrastructure and access to useful data to fuel AI solutions. The assessment of current basic capacity is the first step for AI strategy development. Once this is done, strengthening infrastructure and evaluating the implications of data-sharing processes are key. City leaders should not only initiate and fund the implementation of the necessary infrastructure but should also ensure interoperability and system integration.

5.4. DEVELOP INNOVATIVE REGULATORY TOOLS FOR AI

Regulation is a key tool for cities to direct the development of AI and its interaction in the local environment. Cities can use both soft and hard levers effectively in their jurisdiction.

RECOMMENDATION #9: CREATE AN ENABLING ENVIRONMENT.

While governments on all levels may not always be early adopters of digital innovation, they obviously play a key role in shaping the context of AI in the city.

With an overview of the different sectors, defining the parameters of innovation through local regulation, and with a finger on the pulse of what it is like to live in the city, local authorities have the power to create the ways disruptive technologies will be used to better serve citizens. Local authorities often have a big impact on creating the conditions for AI development and in the city. Building an enabling environment for AI means creating the conditions for responsible AI in the city, beyond building the internal capacity of city governments alone.

Developing AI in the city will require medium- to long-term change. This time horizon is sometimes difficult to discuss and implement when politicians focus on short-term priorities alone (Prins et al., 2021). While current events may serve as a catalyst for digital transformation, every seed needs fertile soil. Local governments can help create that soil.

RECOMMENDATION #10: INTRODUCE LOCAL TECHNOLOGICAL STANDARDS AND CERTIFICATIONS.

Technical standards, explainability standards and ethical standards can be useful regulatory tools. A promising perspective is the design of data-sharing standards which enable the city to use the data collected by private actors as well as facilitating collaborative governance.

The implementation of these standards can be supported by establishing certification systems for those who work with AI and developing the policies that serve to implement them (Prins et al., 2021).

Data standards: Open Mobility Foundation

The Open Mobility Foundation is a non-profit that developed the Open Mobility Standards. It is an "open-source standard that includes real-time reporting through an API" (D'Agostino et al., 2019). It tracks individual mobility using a unique ID, creating a valuable location database. Originally developed in Los Angeles, it now operates in more than 130 cities around the world.

The development is important because the City of Los Angeles uses the standards as a precondition for micro-mobility services; for example, in order for shareable scooters to develop their services, they must use these open data standards, effectively sharing their data with the city. Through the standard, the city benefits from the urban service as well as from the data.

RECOMMENDATION #11: INCORPORATE AND ADAPT AI ASSESSMENT TOOLS.

Evaluation is not a one-time-only process; it happens continuously and re-occurs. Cities need to design procedures with the longer term in mind, so that when things change the city can respond reflexively and adapt.

Different impact assessment tools are being explored as promising methods for accountability across the AI life cycle. Examples include Algorithm Impact Assessments (AIA) and Human Rights Impact Assessments (HRIA), which can form new accountability relationships and governance architectures. Human Rights Impact Assessments are existing methodologies that can be adapted for AI systems. They can help designers and implementers of the system to study its impact through correspondence with the rights-holders (e.g., the citizens in the city) and external stakeholders (Latonero, 2018). Including these mechanisms and regulations works towards reliability, safety and trustworthiness over time. These mechanisms are ways to incorporate the input of a broader array of stakeholders, including auditors, researchers and civil society (Nagitta et al., 2022).

RECOMMENDATION #12: BUILD ON EXISTING MONITORING AND EVALUATION TO OVERSEE AI SYSTEMS AT SEVERAL POINTS IN TIME.

Evaluation must be ongoing, particularly as the AI life cycle has several phases. Monitoring must include both how an AI system is working, but also its impact after deployment.

Build a robust monitoring and evaluation framework for your AI systems. Take existing evaluation frameworks and build on them, connecting with algorithmic auditing and impact assessments. Monitoring should include a skilled interdisciplinary team, using one-, three- and six-year cycles.

It is also important that monitoring frameworks consider the public interest. While existing political processes are founded on representation, it is insufficient to assume that political processes are enough on their own to align AI implementation with public values. A separate mechanism for oversight is required. It can be useful to carry out a reflexive exercise to express which values the city chooses and how to operationalise them (Jameson et al., 2021).

Algorithmic impact assessments

Algorithmic impact assessments (AIAs) estimate the harms caused by an AI to society and offer measures to mitigate those harms. AIAs often look at variables such as the actors, the methods and the setting where algorithms are deployed. Like other impact assessments in other domains, AIAs will not be the answer to all the challenges raised by AI. However, they are currently being developed in an organic process of evolving standards (Metcalf et al., 2021).

AIAs can build on existing human rights impact assessments, but these are often two separate initiatives, and it is recommended to have both.

RECOMMENDATION #13: ADAPT PROCUREMENT PROCESSES.

The vast majority of urban AI will be sourced via procurement. Procurement processes are the city's chance to implement the design strategy. Most cities do not have the in-house capacity to develop robust AI solutions on their own. While developing that capacity is important, most developers writing algorithms work for companies with more financial capacity to invest in higher salaries.

As a result, cities need the capacity to evaluate the AI solutions presented to them during the procurement process. This is a key objective of the risk taxonomy: to enable city administrators to understand and evaluate the risks to be aware of. What questions can you ask when buying an AI solution? One promising methodology for city councils is contractual clauses.

Case study: Amsterdam's procurement clauses

The city of Amsterdam in the Netherlands has been a pioneer in establishing contractual clauses for their public procurement process for algorithms. The clauses focus on technical transparency, procedural transparency and explainability.

The Standard Clauses for Procurement of Trustworthy Algorithmic Systems are openly available and can be freely downloaded from the Amsterdam City Council website (City of Amsterdam Algorithm Register Beta, 2022; Haataja et al., 2021). At the time of this writing, a process is underway to establish standard clauses at the European level.

Case study: Barcelona

Barcelona is considered a pioneer in developing a city strategy for how data, and by extension AI, should be used in the city. The city developed a digital strategy that starts from the vision of putting people first. The idea is value-driven starting from the framing and design phases, beginning with imagining how tech could work differently. To do this, the Barcelona City Council Open Digitisation Plan presents a toolkit known as the Ethical Digital Standards which includes methods, standards, work practices, procurement tools and software standards. Together, these standards set the conditions for working within the city, set the conditions for investment, and create a value-driven environment (Barcelona Digital City, 2016). By feeding back analytic capacities into the city, the city council proactively reverses common trends of extracting data from citizens for profit.

The innovation of public service provision in Barcelona was no accident. Rather, it was enabled by two specific contextual characteristics: first, visionary leadership supported by a political party that came into power, and second, a history of a strong civil society focused on technology as an enabler of power in the city (Monge, 2022b).

5.5. FOSTER CROSS-SECTORAL COLLABORATIONS

Dialogue and collaboration across sectors will be required to develop AI implementation in the city in line with inclusivity and sustainable development goals.

RECOMMENDATION #14: ENCOURAGE LOCAL INNOVATION.

For an effective AI strategy, it is important for the city to consider how to build an environment conducive to communication and partnerships as well as how to invest in the city's capacity to make the most of these opportunities.

For example, when the city provides the space for the regulator to discuss with private actors and small-scale technology entrepreneurs, there is a chance for communication. While these may not be one-time interventions, they create a constructive environment that allows contextually relevant solutions to emerge.

Another method is using urban planning incentives to develop AI locally. Cities can prioritise projects from startups or established companies that serve the public interest. These incentives can take the form of loans, technical assistance, mentorships or even access to land resources.

RECOMMENDATION #15: DEFINE THE TERMS FOR PARTNERSHIPS.

Building partnerships with industry and private actors is often necessary but should be conducted on the city's own terms. The city needs a process to define the metrics and conditions under which it will collaborate. If these terms are socially accepted—particularly when they have emerged as a result of a meaningful participatory process—collaborations may still draw critique, but the process of defining the rules for engagement creates an environment where it's transparent and clear what's being done and why. This clarity creates trust and a space to move forward constructively.

A safe environment for collaboration between civil society and other actors in the ecosystem creates the context of the city, so that when a short-term need or event happens, the city has the resources and connections to adapt and respond. Civil society that focuses on technology innovation can also create innovative, decentralised initiatives for AI and data governance developments. The challenge is that these local initiatives are often unable to scale without the support from government or political parties (Monge, 2022a).

RECOMMENDATION #16: ENGAGE THE PRIVATE SECTOR.

Technological businesses can concentrate the appropriate skills to both develop and manage an AI project efficiently and to carry out the relevant R&D needed for innovation. Social media, telecom operators and online sharing platforms can provide local governments with valuable data concerning city agents and the operation of the city services, if the appropriate sharing mechanisms are developed (see the "data standards" box in section 5.4).

Collaboration across sectors must be envisaged before and beyond procurement to create the conditions for success. Public-private partnerships may be challenging, as public and private organisations often do not share the same objectives or the same timelines. The relatively short-term reasoning of businesses, in line with shareholders' agendas, may lead to very different visions of AI implementation. In particular, their approach to risk management can diverge, as businesses rarely consider the same long-term political risks that a sustainable AI strategy should address. While city politics often involve focusing on short-term priorities, it is important that the city strategy re-emphasise long-term issues (Prins et al., 2021).

New ways of engaging with the private sector can be found as businesses are more and more encouraged to commit to responsible AI. More than 100 leading organisations have joined the Partnership on AI in order to develop an AI to empower humanity. Guiding principles of responsible AI have been published by firms that make available tools for the management and implementation of AI.

RECOMMENDATION #17: ENGAGE PUBLIC RESEARCH.

Public research institutes and universities around the world have a meaningful role to carry out in support of local and national leaders. Their position as independent agents that display no commercial interests make them a very important actor in the public-private relationship (Gasser and Almeida, 2017).

The research sector provides resources that facilitate the development and deployment of AI, including support for the assessment of AI. On a technical level, they can develop specific measures to assess the accuracy and fairness of the outcome. On an implementation level, they can conduct impact assessments. Researchers can provide valuable local knowledge and evidence as a base for policymaking, particularly from multi-dimensional and interdisciplinary perspectives. Social and legal experts can also provide significant input into the framing stage of the AI life cycle.

Research centres are therefore particularly well suited to inducing engagement and inclusion. The resources and capacity nurtured by universities are indispensable for the functioning of both businesses and governments. In that sense, they represent a privileged space for dialogue between private and public entities. Governing organisations should harness these strengths.

RECOMMENDATION #18: ENGAGE WITH CIVIL SOCIETY.

The particular role of NGOs as a link between people and government places them at the forefront of the movement towards responsible AI. They represent important drivers of AI for good; many NGOs propose AI solutions developed or co-developed with other sectors, specifically focused on the public interest. The mission and activities of NGOs require a profound understanding of the context of intervention and of the impacted population. This expert knowledge can be leveraged by city leaders when defining their strategy or implementing AI solutions.

NGOs may also collect data on behalf of marginalised communities or neighbourhoods with which they have worked closely. Furthermore, they can help shape the data collection process by identifying information loopholes, focusing on issues that have not been prioritised by local authorities.

An AI strategy should carefully consider how to protect the space for civil society to operate. Civil society is a powerful voice that can speak for the communities it represents and can hold others accountable for their actions and their impacts on society. City governments may channel the close relationship of the civil society and the public to raise awareness on the opportunities and risks relating to AI. In certain contexts, NGOs have been responsible for building digital capacity within cities by providing technical equipment and encouraging digital literacy for all.

5.6. BUILD HORIZONTAL INTEGRATION

RECOMMENDATION #19: CREATE A MORE INTEGRATED MUNICIPAL STRUCTURE.

New municipal structures or organisations may be a useful tool to carry out the vision for an AI strategy. Integrated organisational structures are one method to integrate silos across an urban municipality (Leslie et al., 2021). They can coordinate policymaking across scale levels or engage with other actors in a co-production.

The approach of creating new, cross-sector, cross-discipline integrated structures with the specific mandate to direct and manage data or AI-focused initiatives has worked particularly well for larger urban agglomerations, such as London's regional planning office and Barcelona's municipal data office. For smaller urban centres, the key is to identify the actor or coalition of actors from the public, private, research and non-profit spheres that can support the AI strategy.

In order to appreciate the benefit of independent regulatory and oversight institutions, expert committees or sectoral regulators, a needs assessment should be carried out (Bulmer, 2019). Depending on the existing capacity of local authorities or the particular jurisdiction context, the implementation of an adequate regulatory landscape may require the implementation of new ad-hoc bodies (United Nations, 2019).

Algorithm registers

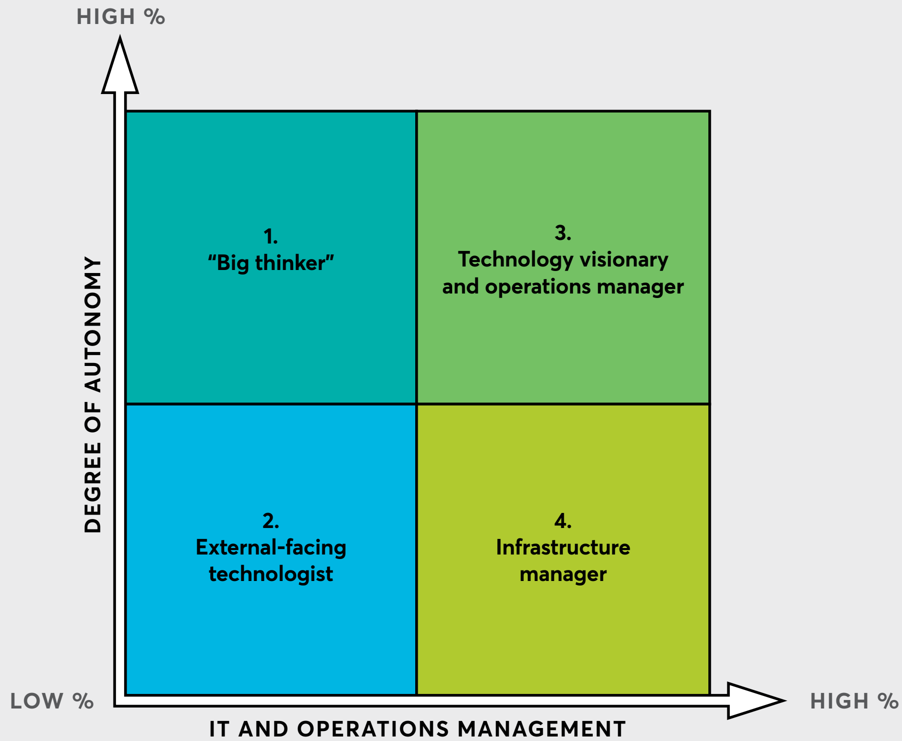
One tool to break down information silos for AI governance is algorithm registers. An algorithm register is an "overview of the artificial intelligence systems and algorithms used by the city" (City of Amsterdam Algorithm Register Beta, 2020), including the reasons they are being used and an explanation of the way they function. Part of the challenge in governing AI is that locally there is often limited understanding of what algorithms are actually in use and what they do. Algorithm registers are a way to address that challenge.

RECOMMENDATION #20: NURTURE INNOVATION LEADERSHIP WITH CTOS AND CIOS.

A very interesting way for local authorities to be proactive on digital innovation is to nurture similar leadership positions as those found in the tech industry: Chief Innovation Officer (CIO) or Chief Technology Officer (CTO). These are terms originating from industry and are relatively new to applications in urban development. CTOs are often at the helm of municipal reorganisation.

Key roles: Chief Technology Officers

Innovation leaders function in a few different ways. These roles can be broadly described as a spectrum between the degree of autonomy and operations management:



1. CTO as "big thinker": In this mode, the CTO is given a lot of leeway to think about long-term development and future approaches.

2. CTO as external-facing technologist: As an external-facing technologist, the CTO concentrates their efforts on collaborating closely with city stakeholders to design and establish digital innovation.

3. CTO as technology visionary and operations manager: This model combines models 1 and 2. Here the CTO is brought in early in the strategy planning process. The CTO is in charge of figuring out how technology may be leveraged to carry out the proposed strategy and then is responsible for executing the plan.

4. CTO as infrastructure manager: The CTO demonstrates operational skills, a clear awareness of technology management, and the ability to oversee a large and diverse team. In this mode, their main goal is to keep the IT department running smoothly, rather than make decisions on technological strategy.

SECTION 6

Conclusion

The field of AI is growing at an unbridled pace. We are increasingly seeing AI systems leaving research settings to be deployed in almost all spheres of human activity. As a result, AI has the potential to profoundly transform the way our societies operate, including by supporting efforts on critical questions such as the climate crisis, public health, education and beyond. However, this ongoing societal transformation entails risks that must be addressed. There is an urgent need to develop responsible AI governance and practices across all scale levels of administrative and political organisations, in both the public and private sectors.

This report provides a general framework on how to deploy AI responsibly in the context of cities and settlements. It offers an overview of the major considerations facing local authorities as they make important decisions on how and when to use AI. The report provides a review of AI governance in urban contexts, an analysis of existing AI applications. It proposes a Risk Assessment Framework that spans the entire AI life cycle and makes a set of recommendations for policy makers to consider when drafting AI strategies. Together, these elements support Mila's commitment to advancing AI for the benefit of all and UN-Habitat's vision of a better quality of life for all in an urbanising world.

While we hope this report is helpful, most of the work lies ahead. Leadership, knowledge and planning will be required for decision-makers to implement AI strategies that are responsible, inclusive and ambitious. While this report provides recommendations to this end, there is no standardised recipe for success, as local contexts must play a pivotal role in designing any AI strategy.

In order to better support decision-makers in this exercise, future work should explore at least three important areas: first, highlight the experiences of non-Western cities implementing AI applications and consider how to support capacity-building in ways that are globally equitable; second, examine how AI can support practical urban planning processes in further detail; and third, develop tools and processes to meaningfully include local populations and civil society organisations all along the AI life cycle.

Finally, we invite feedback on this report. We would love to hear from cities that are actually using this report on what worked, what is helpful, and what should be improved to better respond to their local contexts. This feedback will help us develop our future thinking and provide adapted advice and thought leadership to enable responsible AI across domains and contexts. Together, Mila and UN-Habitat believe that when decision-makers are informed about both the risks and benefits of AI, they are better positioned to use AI as a tool for creating inclusive, safe, resilient and sustainable cities and communities, as well as reducing inequality, discrimination and poverty. This report is our humble contribution in this direction. We hope that decision-makers at all levels of government will use and share this report widely for the betterment of cities and settlements worldwide.

REFERENCES

Abbasi, Maryam, and Ali El Hanandeh (2016). Forecasting municipal solid waste generation using artificial intelligence modelling approaches. *Waste Management*, No. 56 (October), pp. 13–22. https://doi.org/10.1016/j.wasman.2016.05.018

Abdallah, Mohamed, Manar Abu Talib, Sainab Feroz, Qassim Nasir, Hadeer Abdalla, and Bayan Mahfood (2020). Artificial intelligence applications in solid waste management: A systematic research review. *Waste Management*, No. 109 (May), pp. 231–46. https://doi.org/10.1016/j.wasman.2020.04.057

Abioye, Sofiat O., Lukumon O. Oyedele, Lukman Akanbi, Anuoluwapo Ajayi, Juan Manuel Davila Delgado, Muhammad Bilal, Olugbenga O. Akinade, and Ashraf Ahmed (2021). Artificial intelligence in the construction industry: A review of present status, opportunities and future challenges. *Journal of Building Engineering*, No. 44 (December). https://doi.org/10.1016/j.jobe.2021.103299

Aguilar, Diego, Roxana Barrantes, Aileen Agüero, Onkokame Mothobi, and Tharaka Amarasinghe (2020). Future of work in the Global South: Digital labor, new opportunities and challenges. Instituto de Estudios Peruanos. https://afteraccess.net/wp-content/uploads/AfterAccess-Future-of-Work-in-the-Global-South-Digital-Labor-New-Opportunities-and-Challenges-Working-Paper.pdf

Ahmad, Kashif, Majdi Maabreh, Mohamed Ghaly, Khalil Khan, Junaid Qadir, and Ala Al-Fuqaha (2021). Developing future human-centered smart cities: Critical analysis of smart city security, interpretability, and ethical challenges. *arXiv:2012.09110 [cs]* (5 December). http://arxiv.org/abs/2012.09110

Aitken, Rob (2017). "All data is credit data": Constituting the unbanked. *Competition & Change*, vol. 21, No. 4 (1 August), pp. 274–300. https://doi.org/10.1177/1024529417712830

Akanbi, Lukman A., Ahmed O. Oyedele, Lukumon O. Oyedele, and Rafiu O. Salami (2020). Deep learning model for demolition waste prediction in a circular economy. *Journal of Cleaner Production*, No. 274 (November). https://doi.org/10.1016/j.jclepro.2020.122843

Alam, Gulzar, Ihsanullah Ihsanullah, Mu Naushad, and Mika Sillanpää (2022). Applications of artificial intelligence in water treatment for optimization and automation of adsorption processes: Recent advances and prospects. *Chemical Engineering Journal*, No. 427 (January), p. 130011. https://doi.org/10.1016/j.cej.2021.130011

Alan Turing Institute (2018). New collaboration between the Alan Turing Institute and Digital Catapult provides funding for London start-ups and scale-ups to open their data challenges to researchers. News post, 16 August. https://www.turing.ac.uk/news/new-collaboration-between-alan-turing-institute-and-digital-catapult-provides-funding-london

Ala-Pietilä, Pekka, and Nathalie A. Smuha (2021). A framework for global cooperation on artificial intelligence and its governance. In *Reflections on Artificial Intelligence for Humanity*, Bertrand Braunschweig and Malik Ghallab, eds., pp. 237–265. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-69128-8_15

Almeida, Denise, Constantin Shmarko, and Elizabeth Lomas (2021). The ethics of facial recognition technologies, surveillance, and accountability in an age of artificial intelligence: a comparative analysis of US, EU, and UK regulatory frameworks. *AI and Ethics* No. 2, pp. 377–387. https://doi.org/10.1007/s43681-021-00077-w

Almeida, Patricia Gomes Rêgo de, Carlos Denner dos Santos, and Josivania Silva Farias (2021). Artificial intelligence regulation: A framework for governance. *Ethics and Information Technology*, vol. 23, No. 3 (1 September), pp. 505–25. https://doi.org/10.1007/s10676-021-09593-z

Amir Haeri, Maryam, and Katharina Zweig (2020). The crucial role of sensitive attributes in fair classification, 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 2993–3002. https://doi.org/10.1109/SSCI47803.2020.9308585

Amodei, Dario, Chris Olah, Jacob Steinhardt, Paul Christiano, John Schulman, and Dan Mané (2016). Concrete problems in AI safety. *arXiv:1606.06565 [cs]* (25 July). http://arxiv.org/abs/1606.06565

Anttiroiko, Ari-Veikko, Pekka Valkama, and Stephen Bailey (2014). Smart cities in the new service economy: Building platforms for smart services. *AI & Society*, No. 29 (3 November), pp. 323–34. https://doi.org/10.1007/s00146-013-0464-0

As, Imdat, Siddharth Pal, and Prithwish Basu (2018). Artificial intelligence in architecture: Generating conceptual design via deep learning. *International Journal of Architectural Computing*, vol. 16, No. 4 (December), pp. 306–27. https://doi.org/10.1177/1478077118800982

Asaro, Peter (2019). AI ethics in predictive policing: From models of threat to an ethics of care. *IEEE Technology and Society Magazine*, No. 38 (1 June), pp. 40–53. https://doi.org/10.1109/MTS.2019.2915154

Associated Press (2022). Oregon is dropping an artificial intelligence tool used in child welfare system. *NPR*, 2 June. https://www.npr.org/2022/06/02/1102661376/oregon-drops-artificial-intelligence-child-abuse-cases

Babyl (2018). The Rwandan government partners with babyl to deliver the first ever fully digital healthcare service in East Africa using artificial intelligence. Press release, 10 January. https://www.babyl.rw/uploads/press-releases/Press-Release-MoH-babyl-VF.pdf

Barcelona Digital City (2016). Barcelona City Council digital plan: A government measure or open digitisation: Free software and agile development of public administration services. https://ajuntament.barcelona.cat/digital/sites/default/files/LE_MesuradeGovern_EN_9en.pdf

Barcelona Digital City (2021). Barcelona promotes the ethical use of artificial intelligence. Blog, 21 April. https://ajuntament.barcelona.cat/digital/en/blog/barcelona-promotes-the-ethical-use-of-artificial-intelligence

Bastani, Favyen, Songtao He, Sofiane Abbar, Mohammad Alizadeh, Hari Balakrishnan, Sanjay Chawla, and Sam Madden (2018). Machine-assisted map editing. *Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* (November 6), pp. 23–32. https://doi.org/10.1145/3274895.3274927

REFERENCES

Basu, Medha (2017). Exclusive: Hong Kong's vision for artificial intelligence. *GovInsider*, 6 October. https://govinsider.asia/smart-gov/exclusive-hong-kongs-vision-for-artificial-intelligence/

Bekkum, Marvin van, and Frederik Zuiderveen Borgesius (2021). Digital welfare fraud detection and the Dutch SyRI judgment. *European Journal of Social Security*, vol. 23, No. 4, pp. 323–40. https://doi.org/10.1177/13882627211031257

Bench-Capon, Trevor, Michał Araszkiewicz, Kevin Ashley, Katie Atkinson, Floris Bex, Filipe Borges, Daniele Bourcier, et al. (2012). A history of AI and law in 50 papers: 25 years of the International Conference on AI and Law. *Artificial Intelligence and Law* 20, No. No. 3 (September), pp. 215–319. https://doi.org/10.1007/s10506-012-9131-x

Bengio, Y., A. Courville, and P. Vincent (2013). Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, No. 8 (August), pp. 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Berryhill, Jamie, Kevin Kok Heang, Rob Clogher, and Keegan McBride (2019). Hello, world Artificial intelligence and its use in the public sector. OECD Working Papers on Public Governance. https://www.oecd-ilibrary.org/governance/hello-world_726fd39d-en

Bhattacharya, Biswarup, and Abhishek Sinha (2017). Deep fault analysis and subset selection in solar power grids. *arXiv:1711.02810 [cs.LG]*. https://doi.org/10.48550/ARXIV.1711.02810

Bolukbasi, Tolga, Kai-Wei Chang, James Zou, Venkatesh Saligrama, and Adam Kalai (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. *arXiv* (July 21). http://arxiv.org/abs/1607.06520

Borgs, Christian, Ozan Candogan, Jennifer Chayes, Ilan Lobel, and Hamid Nazerzadeh (2014). Optimal multiperiod pricing with service guarantees. *Management Science*, vol. 60, No. 7 (July 1), pp. 1792–1811. https://doi.org/10.1287/mnsc.2013.1839

Bouhedda, Mounir, Sonia Lefnaoui, Samia Rebouh, and Madiha M. Yahoum (2019). Predictive model based on adaptive neuro-fuzzy inference system for estimation of cephalexin adsorption on the octenyl succinic anhydride starch. *Chemometrics and Intelligent Laboratory Systems*, No. 193 (October), p. 103843. https://doi.org/10.1016/j.chemolab.2019.103843

Brandusescu, A., and Reia, J., eds. (2022). *Artificial Intelligence in the City: Building Civic Engagement and Public Trust*. Montreal: Centre for Interdisciplinary Research on Montréal, McGill University. https://www.researchgate.net/publication/362025076_Artificial_Intelligence_in_the_City_Building_Civic_Engagement_and_Public_Trust

BT (n. d.). Intelligent cities case study: Saving money and reducing emissions in Milton Keynes using smart parking. https://www.iot.bt.com/assets/documents/bt-milton-keynes-innovative-parking-case-study.pdf

Bughin, Jacques, and Nicolas Van Zeebroeck (2018). AI adoption: Why a digital base is critical. McKinsey. *McKinsey Quarterly*, 26 July. https://www.mckinsey.com/business-functions/quantumblack/our-insights/artificial-intelligence-why-a-digital-base-is-critical

Bulmer, Eliott (2019). *Independent regulatory and oversight (fourth-branch) institutions*. International Institute for Democracy and Electoral Assistance. https://www.idea.int/publications/catalogue/independent-regulatory-and-oversight-fourth-branch-institutions

Buolamwini, Joy, and Timnit Gebru (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, pp. 77–91. Proceedings on Machine Learning Research. https://proceedings.mlr.press/v81/buolamwini18a.html

Butler, Keith T., Daniel W. Davies, Hugh Cartwright, Olexandr Isayev, and Aron Walsh (2018). Machine learning for molecular and materials science. *Nature*, vol. 559, No. 7715 (July), pp. 547–55. https://doi.org/10.1038/s41586-018-0337-2

Buuren, Arwin van (2009). Knowledge for governance, governance of knowledge: Inclusive knowledge management in collaborative governance processes. *International Public Management Journal*, vol. 12, No. 2, pp. 208–35. https://doi.org/10.1080/10967490902868523

Caliva, Francesco, Fabio Sousa De Ribeiro, Antonios Mylonakis, Christophe Demazirere, Paolo Vinai, Georgios Leontidis, and Stefanos Kollias (2018). A deep learning approach to anomaly detection in nuclear reactors. In *2018 International Joint Conference on Neural Networks (IJCNN)*,Rio de Janeiro: IEEE, pp. 1–8. https://doi.org/10.1109/IJCNN.2018.8489130

Cameron, Felix (2018). Artificial intelligence and proptech. *Data Driven Investor* (1 May). https://medium.datadriveninvestor.com/artificial-intelligence-and-proptech-ca4f419c7735

Cenek, Martin, Rocco Haro, Brandon Sayers, and Jifeng Peng (2018). Climate change and power security: Power load prediction for rural electrical microgrids using long short term memory and artificial neural networks. *Applied Sciences*, vol. 8, No. 5 (9 May), p. 749. https://doi.org/10.3390/app8050749

Chaillou, Stanislas (2022). *Artificial Intelligence and Architecture: From Research to Practice* 1st ed. Boston: De Gruyter.

Chakraborty, S., and A. C. Newton (2011). Climate change, plant diseases and food security: An overview: Climate change and food security. *Plant Pathology*, vol. 60, No. 1 (February), pp. 2–14. https://doi.org/10.1111/j.1365-3059.2010.02411.x

Chakravorti, Bhaskar (2021). How to close the digital divide in the U.S. *Harvard Business Review*, 20 July . https://hbr.org/2021/07/how-to-close-the-digital-divide-in-the-u-s

Chandler, Jenna (2020). The Northridge earthquake exposed flaws in the Getty's construction—and changed how LA builds. *Curbed Los Angeles*, 17 January. https://la.curbed.com/2020/1/17/21068895/los-angeles-earthquake-steel-buildings-getty

REFERENCES

Chandra, Priyanjani, Pratool Bharti, and Michael Papka (2020). A computer vision and AI based solution to determine the change in water level in stream. Poster, SC20: International Conference for High Performance Computing, Networking, Storage, and Analysis. http://sc20.supercomputing.org/proceedings/src_poster/src_poster_pages/spostg110.html

Cho, Peter Jaeho, Karnika Singh, and Jessilyn Dunn (2021). Chapter 9 – Roles of artificial intelligence in wellness, healthy living, and healthy status sensing. In *Artificial Intelligence in Medicine*, Lei Xing, Maryellen L. Giger, and James K. Min, eds. Academic Press, pp. 151–72. https://doi.org/10.1016/B978-0-12-821259-2.00009-0

Chui, Kwok Tai, Miltiadis D. Lytras, Anna Visvizi, and Akila Sarirete (2021). Chapter 16 – An overview of artificial intelligence and big data analytics for smart healthcare: Requirements, applications, and challenges. In *Artificial Intelligence and Big Data Analytics for Smart Healthcare*, Miltiadis D. Lytras, Akila Sarirete, Anna Visvizi, and Kwok Tai Chui, eds. Academic Press, pp. 243–54. https://doi.org/10.1016/B978-0-12-822060-3.00015-2

Chui, Michael, Martin Harryson, James Manyika, Roger Roberts, Rita Chung, Ashley van Heteren, and Pieter Nel (2018). Notes from the AI frontier: Applying AI for social good. Discussion paper. McKinsey Global Institute, December. https://www.mckinsey.com/~/media/mckinsey/featured%20insights/artificial%20intelligence/applying%20artificial%20intelligence%20for%20social%20good/mgi-applying-ai-for-social-good-discussion-paper-dec-2018.ashx

City of Amsterdam Algorithm Register Beta (2020). What is the Algorithm Register? https://algoritmeregister.amsterdam.nl/en/ai-register/

Clinton, Nicholas, and Peng Gong (2013). MODIS detected surface urban heat islands and sinks: Global locations and controls. *Remote Sensing of Environment*, No. 134 (July), pp. 294–304. https://doi.org/10.1016/j.rse.2013.03.008

Clutton-Brock, Peter, David Rolnick, Priya L. Donti, Lynn H. Kaack, Tegan Maharaj, Alexandra Luccioni, and Hari Prasanna Das (2021). Climate change and AI: Recommendations for government action. Report. GPAI (Global Partnership on AI). https://www.gpai.ai/projects/climate-change-and-ai.pdf

Crawford, Kate (2013). Think again: Big data. *Foreign Policy* (blog), 10 May. https://foreignpolicy.com/2013/05/10/think-again-big-data/

Crawford, Kate (2021). *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence*. New Haven, CT: Yale University Press. https://yalebooks.yale.edu/9780300264630/atlas-of-ai

Criado-Perez, Caroline (2019). *Invisible Women: Data Bias in a World Designed for Men*. New York: Abrams Press.

D'Agostino, M., Pellaton, P., and Brown, A. (2019). Mobility data sharing: Challenges and policy recommendations. *UC Davis: Institute of Transportation Studies*. https://escholarship.org/uc/item/47p885q8

Dai, Xiaoqing, Lijun Sun, and Yanyan Xu (2018). Short-term origin-destination based metro flow prediction with probabilistic model selection approach. *Journal of Advanced Transportation* (26 June), pp. 1–15. https://doi.org/10.1155/2018/5942763

Das, Utpal Kumar, Kok Soon Tey, Mehdi Seyedmahmoudian, Saad Mekhilef, Moh Yamani Idna Idris, Willem Van Deventer, Bend Horan, and Alex Stojcevski (2018). Forecasting of photovoltaic power generation and model optimization: A review. *Renewable and Sustainable Energy Reviews*, No. 81 (January), pp. 912–28. https://doi.org/10.1016/j.rser.2017.08.017

Data Society (2021). Assembling accountability: Algorithmic impact assessment for the public interest (June). https://datasociety.net/wp-content/uploads/2021/06/Assembling-Accountability.pdf

Daughton, Ashlynn R., and Michael J. Paul (2019). Identifying protective health behaviors on Twitter: Observational study of travel advisories and Zika virus. *Journal of Medical Internet Research*, vol. 21, No. 5), p. e13090. https://doi.org/10.2196/13090

Davenport, Thomas H., and D. J. Patil (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 1 October. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Davenport, Thomas, and Ravi Kalakota (2019). The potential for artificial intelligence in healthcare. *Future Healthcare Journal*, vol. 6, No. 2 (June), pp. 94–98. https://doi.org/10.7861/futurehosp.6-2-94

Decuyper, Adeline, Alex Rutherford, Amit Wadhwa, Jean-Martin Bauer, Gautier Krings, Thoralf Gutierrez, Vincent D. Blondel, and Miguel A. Luengo-Oroz (2014). Estimating food consumption and poverty indices with mobile phone data. *arXiv:1412.2595 [physics]*, 22 November. http://arxiv.org/abs/1412.2595

Dhinakaran, Aparna (2021). Overcoming AI's transparency paradox. *Forbes*, 10 September. https://www.forbes.com/sites/aparnadhinakaran/2021/09/10/overcoming-ais-transparency-paradox/?sh=76d9fde34b77

Dietvorst, Berkeley J., Joseph P. Simmons, and Cade Massey (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, vol. 144, No. 1, pp. 114–26. https://doi.org/10.1037/xge0000033

Dignum, Virginia (2022). Responsible artificial intelligence – from principles to practice. *ACM SIGIR Forum*, vol. 56, No. 1 (June). https://arxiv.org/pdf/2205.10785.pdf

Dijk, Jan A.G.M. van (2006). Digital divide research, achievements and shortcomings. *Poetics*, vol. 34, No. 4–5 (August), pp. 221–35. https://doi.org/10.1016/j.poetic.2006.05.004

Ditzler, Gregory, Manuel Roveri, Cesare Alippi, and Robi Polikar (2015). Learning in nonstationary environments: A survey. *Computational Intelligence Magazine, IEEE* 10 (1 November), pp. 12–25. https://doi.org/10.1109/MCI.2015.2471196

Dobbe, Roel, David Fridovich-Keil, and Claire Tomlin (2017). Fully decentralized policies for multi-agent systems: An information theoretic approach. *arXiv:1707.06334 [nlin]* (29 July). http://arxiv.org/abs/1707.06334

Dobbe, Roel, Oscar Sondermeijer, David Fridovich-Keil, Daniel Arnold, Duncan Callaway, and Claire Tomlin (2018). Towards distributed energy services: Decentralizing optimal power flow with machine learning. *arXiv:1806.06790 [cs.LG]*. https://doi.org/10.48550/ARXIV.1806.06790

Donnot, Benjamin, Isabelle Guyon, Marc Schoenauer, Patrick Panciatici, and Antoine Marot (2017). Introducing machine learning for power system operation support. *arXiv:1709.09527 [cs, stat]* (27 September). http://arxiv.org/abs/1709.09527

REFERENCES

Donti, Priya L., Yajing Liu, Andreas J. Schmitt, Andrey Bernstein, Rui Yang, and Yingchen Zhang (2018). Matrix completion for low-observability voltage estimation. *arXiv:1801.09799 [math.OC]*. https://doi.org/10.48550/ARXIV.1801.09799

Doshi, Jigar, Saikat Basu, and Guan Pang (2018). From satellite imagery to disaster insights. *arXiv:1812.07033 [cs]* (17 December). http://arxiv.org/abs/1812.07033

Doshi-Velez, Finale, and Been Kim (2017). Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608 [cs, stat]* (2 March). http://arxiv.org/abs/1702.08608

Draetta, Laura, and Valerie Fernandez (2021). The Alicem app: A controversial digital authentication system. I'MTech (blog), 3 February. https://imtech.imt.fr/en/2021/02/03/the-alicem-app-a-controversial-digital-authentication-system/

Drozdowski, Pawel, Christian Rathgeb, Antitza Dantcheva, Naser Damer, and Christophe Busch (2020). Demographic bias in biometrics: A survey on an emerging challenge. *Institute of Electrical and Electronics Engineers (IEEE)* (5 March).

Dwork, Cynthia, and Christina Ilvento (2018). Fairness under composition. *arXiv:1806.06122 [cs.LG]* (15 June). https://arxiv.org/abs/1806.06122

Dwork, Cynthia, and Marthe Louise Minow (2022). Distrust of artificial intelligence: Sources & responses from computer science & law. *Daedalus*, vol. 151, No. 2, pp. 309–321.(spring). https://doi.org/10.1162/DAED_a_01918

Eastin, Matthew S., and Robert LaRose (2000). Internet self-efficacy and the psychology of the digital divide. *Journal of Computer-Mediated Communication*, vol. 6, No. 1 (23 June), pp. 0–0. https://doi.org/10.1111/j.1083-6101.2000.tb00110.x

El Hanandeh, Ali, Zainab Mahdi, and M. S. Imtiaz (2021). Modelling of the adsorption of Pb, Cu and Ni ions from single and multi-component aqueous solutions by date seed derived biochar: Comparison of six machine learning approaches. *Environmental Research*, No. 192 (January. https://doi.org/10.1016/j.envres.2020.110338

Ellman, Douglas (Douglas Austin) (2015). The reference electrification model: A computer model for planning rural electricity access. Thesis, Massachusetts Institute of Technology. https://dspace.mit.edu/handle/1721.1/98551

Enderlein, Henrik, Sonja Wälti, and Michael Zürn (2010). *Handbook on Multi-Level Governance*. Cheltenham: Elgar Publishing. https://opus4.kobv.de/opus4-hsog/frontdoor/index/index/docId/298

Engler, Alex (2021). Enrollment algorithms are contributing to the crises of higher education. *Brookings* (blog), 14 September. https://www.brookings.edu/research/enrollment-algorithms-are-contributing-to-the-crises-of-higher-education/

European Data Protection Board (2021). Dutch DPA issues formal warning to a supermarket for its use of facial recognition technology. Blog, 26 January. https://edpb.europa.eu/news/national-news/2021/dutch-dpa-issues-formal-warning-supermarket-its-use-facial-recognition_en

Ewijk, Edith van, and I. S. A. Baud (2009). 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, pp. 218–26. https://doi.org/10.1016/j.habitatint.2008.10.014

Faggella, Daniel (2015). Can abuse of AI agthe shape the future of human computer interaction? Singularity blog, 13 November. https://www.singularityweblog.com/can-abuse-of-ai-agents-shape-the-future-of-human-computer-interaction/

Falco, Gregory, Ben Shneiderman, Julia Badger, Ryan Carrier, Anton Dahbura, David Danks, Martin Eling, et al. (2021). Governing AI safety through independent audits. *Nature Machine Intelligence*, No. 3 (1 July), pp. 566–71. https://doi.org/10.1038/s42256-021-00370-7

Floridi, Luciano (2018). Soft ethics and the governance of the digital. *Philosophy & Technology*, vol. 31, No. 1 (1 March), pp. 1–8. https://doi.org/10.1007/s13347-018-0303-9

Floridi, Luciano (2019). Establishing the rules for building trustworthy AI. *Nature Machine Intelligence*, No. 1, pp. 261–262 (7 May). https://philpapers.org/rec/FLOETR

Foley, Aoife M., Paul G. Leahy, Antonino Marvuglia, and Eamon J. McKeogh (2012). Current methods and advances in forecasting of wind power generation. *Renewable Energy*, vol. 37, No. 1 (January), pp. 1–8. https://doi.org/10.1016/j.renene.2011.05.033

Fujimura, Koji, Atsuto Seko, Yukinori Koyama, Akihide Kuwabara, Ippei Kishida, Kazuki Shitara, Craig A. J. Fisher, Hiroki Moriwake, and Isao Tanaka (2013). Accelerated materials design of lithium superionic conductors based on first-principles calculations and machine learning algorithms. *Advanced Energy Materials*, vol. 3, No. 8 (August), pp. 980–85. https://doi.org/10.1002/aenm.201300060

Gabriel, Iason (2020). Artificial Intelligence, Values and Alignment. *Minds and Machines*, No. 30 (October), pp. 411–437. https://doi.org/10.1007/s11023-020-09539-2

Galea, Sandro, and David Vlahov (2005). Urban health: Evidence, challenges, and directions. *Annual Review of Public Health*, vol. 26, No. 1), pp. 341–65. https://doi.org/10.1146/annurev.publhealth.26.021304.144708

Gandhi, Tarak, and Mohan Manubhai Trivedi (2008). Computer vision and machine learning for enhancing pedestrian safety. In *Computational Intelligence in Automotive Applications*, D. Prokhorov, ed. Berlin, Heidelberg: Springer, pp. 59–77. https://doi.org/10.1007/978-3-540-79257-4_4

Gar Alalm, Mohamed, and Mahmoud Nasr (2018). Artificial intelligence, regression model, and cost estimation for removal of chlorothalonil pesticide by activated carbon prepared from casuarina charcoal. *Sustainable Environment Research*, vol. 28, No. 3 (May), pp. 101–10. https://doi.org/10.1016/j.serj.2018.01.003

Garg, Nikhil, Londa Schiebinger, Dan Jurafsky, and James Zou (2018). Word embeddings quantify 100 years of gender and ethnic stereotypes. *Proceedings of the National Academy of Sciences*, vol. 115, No. 16 (17 April), pp. E3635–44. https://doi.org/10.1073/pnas.1720347115

Garvie, Claire (2019). Garbage in, garbage out: Face recognition on flawed data. Georgetown Law Center on Privacy & Technology (16 May). https://www.flawedfacedata.com

REFERENCES

Gasser, Urs and Virgilio A. F. Almeida (2017). A layered model for AI governance. *IEEE Internet Computing*, vol. 21, No. 6 (November), pp. 58–62. doi:10.1109/mic.2017.4180835

Gastaldi, Massimiliano, Riccardo Rossi, Gregorio Gecchele, and Luca Della Lucia (2013). Annual average daily traffic estimation from seasonal traffic counts. *Procedia - Social and Behavioral Sciences*, No. 87 (October), pp. 279–91. https://doi.org/10.1016/j.sbspro.2013.10.610

Gebru, Timnit, Jamie Morgenstern, Briana Vecchione, Jennifer Wortman Vaughan, Hanna Wallach, Hal Daumé Iii, and Kate Crawford (2021). Datasheets for datasets. *Communications of the ACM*, vol. 64, No. 12 (December), pp. 86–92. https://doi.org/10.1145/3458723

Ghaemi, Mohammad Sajjad, Bruno Agard, Martin Trépanier, and Vahid Partovi Nia (2017). A visual segmentation method for temporal smart card data. *Transportmetrica A: Transport Science*, vol. 13, No. 5 (28 May), pp. 381–404. https://doi.org/10.1080/23249935.2016.1273273

Ghanem, Ahmed, Mohammed Elhenawy, Mohammed Almannaa, Huthaifa I. Ashqar, and Hesham A. Rakha (2017). Bike share travel time modeling: San Francisco Bay Area case study. In *2017 5th IEEE International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, pp. 586–91. https://doi.org/10.1109/MTITS.2017.8005582

Glielmo, Luigi, Stefania Santini, and Gabriele Serra (1999). A two-time-scale infinite-adsorption model of three-way catalytic converters. In *Proceedings of the 1999 American Control Conference (Cat. No. 99CH36251)*, vol. 4, pp. 2683–87. San Diego: IEEE. https://doi.org/10.1109/ACC.1999.786557

Godoy, Franklin Valdebenito, Nuria Hartmann, and Janina Hanswillemenke (2021). APEC case study: Best practices of smart cities in a digital age. Asia-Pacific Economic Cooperation (APEC) SOM Steering Committee on Economic and Technical Cooperation. https://www.apec.org/publications/2021/08/best-practices-of-smart-cities-in-the-digital-age

Gómez Mont, Constanza, Claudia May Del Pozo, Cristina Martínez Pinto, and Ana Victoria Martín del Campo Alcocer (2020). La inteligencia artificial al servicio del bien social en América Latina y el Caribe: Panorámica regional e instantáneas de doce países. Report, Inter-American Development Bank (May). https://doi.org/10.18235/0002393

Gómez-Bombarelli, Rafael, Jennifer N. Wei, David Duvenaud, José Miguel Hernández-Lobato, Benjamín Sánchez-Lengeling, Dennis Sheberla, Jorge Aguilera-Iparraguirre, Timothy D. Hirzel, Ryan P. Adams, and Alán Aspuru-Guzik (2018). Automatic chemical design using a data-driven continuous representation of molecules. *ACS Central Science*, vol. 4, No. 2 (28 February), pp. 268–76. https://doi.org/10.1021/acscentsci.7b00572

González Perea, Rafael, Emilio Camacho Poyato, Pilar Montesinos, and Juan Antonio Rodríguez Díaz (2019). Optimisation of water demand forecasting by artificial intelligence with short data sets. *Biosystems Engineering*, No. 177 (January), pp. 59–66. https://doi.org/10.1016/j.biosystemseng.2018.03.011

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville (2016). *Deep Learning: Adaptive Computation and Machine Learning*. Cambridge: The MIT Press.

Greenfield, Adam (2013). *Against the Smart City: Part I of The City Is Here for You to Use*. New York City: Do Projects.

Gunasekeran, Dinesh Visva, Rachel Marjorie Wei Wen Tseng, Yih-Chung Tham, and Tien Yin Wong (2021). Applications of digital health for public health responses to COVID-19: A systematic scoping review of artificial intelligence, telehealth and related technologies. *npj Digital Medicine*, vol. 4, No. 1 (26 February), p. 40. https://doi.org/10.1038/s41746-021-00412-9

Gupta, Ritwik, Richard Hosfelt, Sandra Sajeev, Nirav Patel, Bryce Goodman, Jigar Doshi, Eric Heim, Howie Choset, and Matthew Gaston (2019). xBD: A dataset for assessing building damage from satellite imagery. *arXiv:1911.09296 [cs]* (21 November). http://arxiv.org/abs/1911.09296

Haataja, Meeri, Linda van de Fliert, and Pasi Rautio (2020). Public AI registers: Realising AI transparency and civic participation in government use of AI. White paper. https://algoritmeregister.amsterdam.nl/wp-content/uploads/White-Paper.pdf

Hagenauer, Julian, and Marco Helbich (2017). A comparative study of machine learning classifiers for modeling travel mode choice. *Expert Systems with Applications*, No. 78 (July), pp. 273–82. https://doi.org/10.1016/j.eswa.2017.01.057

Hansen, Terry, and Chia-Jiu Wang (2005). Support vector based battery state of charge estimator. *Journal of Power Sources*, vol. 141, No. 2 (March), pp. 351–58. https://doi.org/10.1016/j.jpowsour.2004.09.020

Hao, Karen (2019). AI is sending people to jail and getting it wrong. *MIT Technology Review* 21 January . https://www.technologyreview.com/2019/01/21/137783/algorithms-criminal-justice-ai/

Hao, Karen (2020). Live facial recognition is tracking kids suspected of being criminals. *MIT Technology Review*, 10 September. https://www.technologyreview.com/2020/10/09/1009992/live-facial-recognition-is-tracking-kids-suspected-of-crime/

Heemstra, F. J. (1992). Software cost estimation. *Information and Software Technology*, vol. 34, No. 10 (October), pp. 627–39. https://doi.org/10.1016/0950-5849(92)90068-Z

Ho, Hung Chak, Anders Knudby, Paul Sirovyak, Yongming Xu, Matus Hodul, and Sarah B. Henderson (2014). Mapping maximum urban air temperature on hot summer days. *Remote Sensing of Environment*, No. 154 (November), pp. 38–45. https://doi.org/10.1016/j.rse.2014.08.012

Howard, Philip N., and Samuel C. Woolley (2018). *Computational Propaganda: Political Parties, Politicians, and Political Manipulation on Social Media*. Oxford: Oxford University Press.

Huang, Jeffrey, Mikhael Johanes, Frederick Chando Kim, Christina Doumpioti, and Georg-Christoph Holz (2021). On GANs, NLP and architecture: Combining human and machine intelligences for the generation and evaluation of meaningful designs. *Technology|Architecture + Design*, vol. 5, No. 2 (3 July), pp. 207–24. https://doi.org/10.1080/24751448.2021.1967060

Hussain, F., R. Hussain, and E. Hossain (2021). Explainable artificial intelligence (XAI): An engineering perspective. *arXiv:2101.03613 [cs]* (10 January). http://arxiv.org/abs/2101.03613

REFERENCES

International Telecommunication Union (ITU) (2020). Frontier technologies to protect the environment and tackle climate change. Report. https://www.itu.int/en/action/environment-and-climate-change/Documents/frontier-technologies-to-protect-the-environment-and-tackle-climate-change.pdf

Jacquillat, Alexandre, and Amedeo R. Odoni (2018). A roadmap toward airport demand and capacity management. *Transportation Research Part A: Policy and Practice*, No. 114 (August), pp. 168–85. https://doi.org/10.1016/j.tra.2017.09.027

Jahnke, Art (2018). Are computer-aided decisions actually fair? *The Brink*, 14 December. https://www.bu.edu/articles/2018/algorithmic-fairness/

Jain, Ashu, and Lindell E. Ormsbee (2002). Short-term water demand forecast modeling techniques: Conventional methods versus AI. *Journal - American Water Works Association*, vol. 94, No. 7 (July), pp. 64–72. https://doi.org/10.1002/j.1551-8833.2002.tb09507.x

Jameson, Shazade, Linnet Taylor, and Merel Noorman (2021). Data governance clinics: A new approach to public-interest technology in cities. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, 1 September. https://doi.org/10.2139/ssrn.3880961

Janssen, Marijn, Martijn Hartog, Ricardo Matheus, Aaron Yi Ding, and George Kuk (2022). Will algorithms blind people? The effect of explainable AI and decision-makers' experience on AI-supported decision-making in government. *Social Science Computer Review*, vol. 40, No. 2), pp. 478–93. https://doi.org/10.1177/0894439320980118

Jiang, Huaiguang, and Yingchen Zhang (2016). Short-term distribution system state forecast based on optimal synchrophasor sensor placement and extreme learning machine. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*, pp. 1–5. Boston: IEEE. https://doi.org/10.1109/PESGM.2016.7741933

J. P. Morgan Research (2021). How long will the chip shortage last? https://www.jpmorgan.com/insights/research/supply-chain-chip-shortage

Kaack, Lynn H., George H. Chen, and M. Granger Morgan (2019). Truck traffic monitoring with satellite images. *arXiv:1907.07660 [cs]* (17 July). http://arxiv.org/abs/1907.07660

Kalogirou, Soteris A. (2003). Artificial intelligence for the modeling and control of combustion processes: A review. *Progress in Energy and Combustion Science*, vol. 29, No. 6 (January), pp. 515–66. https://doi.org/10.1016/S0360-1285(03)00058-3

Kassens-Noor, Eva, and Arend Hintze (2020). Cities of the future? The potential impact of artificial intelligence. *AI*, vol, 1, No. 2, pp. 192–197. https://doi.org/10.3390/ai1020012

Khajeh-Hosseini, Ali, David Greenwood, James W. Smith, and Ian Sommerville (2010). The cloud adoption toolkit: Supporting cloud adoption decisions in the enterprise. *arXiv:1008.1900 [cs.DC]*. https://doi.org/10.48550/ARXIV.1008.1900

Kim, Jaewoo, Meeyoung Cha, and Jong Gun Lee (2017). Nowcasting commodity prices using social media. *PeerJ Computer Science*, No. 3 (31 July), p. e126. https://doi.org/10.7717/peerj-cs.126

Kiran, B. Ravi, Ibrahim Sobh, Victor Talpaert, Patrick Mannion, Ahmad A. Al Sallab, Senthil Yogamani, and Patrick Pérez (2021). Deep reinforcement learning for autonomous driving: A survey. *IEEE Transactions on Intelligent Transportation Systems*, vol. 23, No. 6, pp. 1–18. https://doi.org/10.1109/TITS.2021.3054625

Kitchin, Rob (2016). The ethics of smart cities and urban science. *Philosophical Transactions of the Royal Society A*, vol. 374, No. 2083 (28 December), p. 20160115. https://doi.org/10.1098/rsta.2016.0115

Korinek, Anton, Martin Schindler, and Joseph Stiglitz (2021). Technological progress, artificial intelligence, and inclusive growth. *IMF Working Papers* No. 166 (11 June). https://doi.org/10.5089/9781513583280.001.A001

Kornweitz, Aritz (2021). A new AI lexicon: Function creep. Medium, 8 April. https://medium.com/a-new-ai-lexicon/a-new-ai-lexicon-function-creep-1c20834fab4a

Korteling, J. E. (Hans)., G. C. van de Boer-Visschedijk, R. A. M. Blankendaal, R. C. Boonekamp, and A. R. Eikelboom (2021). Human versus artificial intelligence. *Frontiers in Artificial Intelligence*, No. 4. https://doi.org/10.3389/frai.2021.622364

Kosinski, Michal, David Stillwell, and Thore Graepel (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, vol. 110, No. 15 (9 April), pp. 5802–5. https://doi.org/10.1073/pnas.1218772110

Krieger, Nancy, Pamela Waterman, Jarvis Chen, Mah-Jabeen Soobader, S. V. Subramanian, and Rosa Carson (2003). Zip code caveat: Bias due to spatiotemporal mismatches between zip codes and US census-defined geographic areas. *American Journal of Public Health*, vol. 92, No. 7 (1 July), pp. 1100–1102. https://doi.org/10.2105/ajph.92.7.1100

Krile, Robert, Fred Todt, and Jeremy Schroeder (2015). Assessing roadway traffic count duration and frequency impacts on annual average daily traffic estimation: Assessing accuracy issues related to short-term count durations. TRID: the TRIS and ITRD database (October). https://trid.trb.org/view/1442739

Krishna, Kalpesh, Gaurav Singh Tomar, Ankur P. Parikh, Nicolas Papernot, and Mohit Iyyer (2020). Thieves on Sesame Street! Model extraction of BERT-based APIs. *arXiv:1910.12366 [cs]* (12 October). http://arxiv.org/abs/1910.12366

Lao, David (2020). Clearview AI: When can companies use facial recognition data? Global News, 3 June. https://globalnews.ca/news/6621410/clearview-ai-canada-privacy-data/

Larson, Jeff, Surya Mattu, Lauren Kirchner, and Julia Angwin (2016). How we analyzed the COMPAS recidivism algorithm. ProPublica, 23 May. https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm?token=BqO_ITYNAKmQwhj7daSusnn7aJDGaTWE

Larsson, Stefan, and Fredrik Heintz (2020). Transparency in artificial intelligence. *Internet Policy Review*, vol. 9, No. 2 (5 May). https://policyreview.info/concepts/transparency-artificial-intelligence

Latonero, Mark (2018). Governing artificial intelligence: Upholding human rights & dignity. Report. Data & Society, 10 October. https://datasociety.net/library/governing-artificial-intelligence/

Lattimore, Finn, Simon O'Callaghan, Zoe Paelologos, Alistair Reid, Edward Santow, Holli Sargeant, and Andrew Thomsen (2020). Using artificial intelligence to make decisions: Addressing the problem of algorithmic bias. Australian Human Rights Commission (24 November). https://humanrights.gov.au/our-work/rights-and-freedoms/publications/using-artificial-intelligence-make-decisions-addressing

REFERENCES

Lee, Hanbong, Waqar Malik, Bo Zhang, Balaji Nagarajan, and Yoon C. Jung (2015). Taxi time prediction at Charlotte Airport using fast-time simulation and machine learning techniques. In *15th AIAA Aviation Technology, Integration, and Operations Conference*. Dallas, TX: American Institute of Aeronautics and Astronautics. https://doi.org/10.2514/6.2015-2272

Lee, Min Kyung, Daniel Kusbit, Evan Metsky, Laura Dabbish (2015). Working with machines: The impact of algorithmic and data-driven management on human workers. *CHI '15: Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (April) pp. 1603–1612. https://doi.org/10.1145/2702123.2702548

Leslie, David (2019). Understanding artificial intelligence ethics and safety. *arXiv:1906.05684 [cs, stat]* (11 June). https://doi.org/10.5281/zenodo.3240529

Leslie, David, Christopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell, Morgan Briggs, and Tim Clement-Jones (2021). Artificial intelligence, human rights, democracy, and the rule of law. The Alan Turing Institute (June). https://rm.coe.int/primer-en-new-cover-pages-coe-english-compressed-2754-7186-0228-v-1/1680a2fd4a

Leszczynski, Agnieszka (2020). Glitchy vignettes of platform urbanism. *Environment and Planning D: Society and Space*, vol. 38, No. 2 (1 April), pp. 189–208. https://doi.org/10.1177/0263775819878721

Leung, Hareton, and Zhang Fan (2002). Software cost estimation. In *Handbook of Software Engineering and Knowledge Engineering, Volume II: Emerging Technologies*. S. K. Chang, ed. Hackensack, NJ: World Scientific, pp. 307–24. https://doi.org/10.1142/9789812389701_0014

Li, Bo, Peng Qi, Bo Liu, Shuai Di, Jingen Liu, Jiquan Pei, Jinfeng Yi, and Bowen Zhou (2021). Trustworthy AI: From principles to practices. *arXiv:2110.01167 [cs.AI]* (4 October). https://arxiv.org/abs/2110.01167

Licheva, Veronika (2018). Commotion at the Piet Hein Tunnel as misguided tourists cycle right through it. *DutchReview* (blog), 2 August. https://dutchreview.com/news/weird/commotion-at-the-piet-hein-tunnel-as-misguided-tourists-cycle-right-through-it/

Lin, Angela, and Nan-Chou Chen (2012). Cloud computing as an innovation: Perception, attitude, and adoption. *International Journal of Information Management*, vol. 32, No. 6 (December), pp. 533–40. https://doi.org/10.1016/j.ijinfomgt.2012.04.001

Lin, Anthony L., William C. Chen, and Julian C. Hong (2021). Chapter 8 – Electronic health record data mining for artificial intelligence healthcare. In *Artificial Intelligence in Medicine*, Lei Xing, Maryellen L. Giger, and James K. Min, eds. Academic Press, pp. 133–50. https://doi.org/10.1016/B978-0-12-821259-2.00008-9

Lindgren, S. (2020). Data Theory: Interpretive Sociology and Computational Methods. John Wiley & Sons.

Lipton, Zachary (2016). The mythos of model interpretability. *arXiv:1606.03490 [cs.LG]* (10 June). https://arxiv.org/abs/1606.03490

Liu, Lydia (2020). When bias begets bias: A source of negative feedback loops in AI systems. Microsoft Research Blog, 21 January. https://www.microsoft.com/en-us/research/blog/when-bias-begets-bias-a-source-of-negative-feedback-loops-in-ai-systems/

Liu, Yue, Tianlu Zhao, Wangwei Ju, and Siqi Shi (2017). Materials discovery and design using machine learning. *Journal of Materiomics*, vol. 3, No. 3 (September), pp. 159–77. https://doi.org/10.1016/j.jmat.2017.08.002

Lu, Jie, Anjin Liu, Yiliao Song, and Guangquan Zhang (2020). Data-driven decision support under concept drift in streamed big data. *Complex & Intelligent Systems*, vol. 6, No. 1 (1 April), pp. 157–63. https://doi.org/10.1007/s40747-019-00124-4

Luckin, Rose, Wayne Holmes, Mark Griffiths, and Laurie B. Forcier (2016). *Intelligence Unleashed: An Argument for AI in Education*. Pearson Education, London.

Madiega, Tambiama André, and Hendrik Alexander Mildebrath (2021). Regulating facial recognition in the EU. European Parliamentary Research Service. https://www.europarl.europa.eu/RegData/etudes/IDAN/2021/698021/EPRS_IDA(2021)698021_EN.pdf

Malof, Jordan M., Kyle Bradbury, Leslie M. Collins, and Richard G. Newell (2016). Automatic detection of solar photovoltaic arrays in high resolution aerial imagery. *Applied Energy*, No. 183 (December), pp. 229–40. https://doi.org/10.1016/j.apenergy.2016.08.191

Manley, Ed, Chen Zhong, and Michael Batty (2018). Spatiotemporal variation in travel regularity through transit user profiling. *Transportation*, vol. 45, No. 3 (May), pp. 703–32. https://doi.org/10.1007/s11116-016-9747-x

Mäntymäki, Matti, Matti Minkkinen, Teemu Birkstedt, and Mika Viljanen (2022). Defining organizational AI governance. *AI Ethics* (24 February). https://doi.org/10.1007/s43681-022-00143-x

Marr, Bernard (2022). The dangers of not aligning artificial intelligence with human values. *Forbes*, (4 January). https://www.forbes.com/sites/bernardmarr/2022/04/01/the-dangers-of-not-aligning-artificial-intelligence-with-human-values/

Mathe, Johan, Nina Miolane, Nicolas Sebastien, and Jeremie Lequeux (2019). PVNet: A LRCN architecture for spatio-temporal photovoltaic powerforecasting from numerical weather prediction. *arXiv:1902.01453 [cs.LG]* (4 February). https://doi.org/10.48550/ARXIV.1902.01453

Martinho-Truswell, Emma (2018). How AI could help the public sector. *Harvard Business Review* (January). https://hbr.org/2018/01/how-ai-could-help-the-public-sector

Mazloumi, Ehsan, Geoff Rose, Graham Currie, and Sara Moridpour (2011). Prediction intervals to account for uncertainties in neural network predictions: Methodology and application in bus travel time prediction. *Engineering Applications of Artificial Intelligence*, vol. 24, No. 3 (April), pp. 534–42. https://doi.org/10.1016/j.engappai.2010.11.004

Mazzolin, Robert (2020). Artificial intelligence and keeping humans "in the loop." Centre for International Governance Innovation (blog), 23 November. https://www.cigionline.org/articles/artificial-intelligence-and-keeping-humans-loop/

McCarthy, John (2007). What is AI? http://faculty.otterbein.edu/dstucki/inst4200/whatisai.pdf

REFERENCES

McCrimmon, Ryan, and Martin Matishak (2021). Cyberattack on food supply followed years of warnings. *Politico*, 6 May. https://www.politico.com/news/2021/06/05/how-ransomware-hackers-came-for-americans-beef-491936

Mehmood, Hassan, Panos Kostakos, Marta Cortes, Theodoros Anagnostopoulos, Susanna Pirttikangas, and Ekaterina Gilman (2021). Concept drift adaptation techniques in distributed environment for real-world data streams. *Smart Cities*, vol. 4, No. 1 (March), pp. 349–71. https://doi.org/10.3390/smartcities4010021

Mehr, Hila (2017). Artificial intelligence for citizen services and government. Ash Center for Democratic Governance and Innovation. https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

Mehrabi, Ninareh, Fred Morstatter, Nripstuta Saxena, Kristina Lerman, and Aram Galstyan (2019). A survey on bias and fairness in machine learning. *arXiv:1908.09635 [cs.LG]* (23 August). https://arxiv.org/abs/1908.09635

Mehrotra, Dhruv, Suya Mattu, Annie Gilbertson, and Aaron Sankin (2021). How we determined predictive policing software disproportionately targeted low-income, Black, and Latino neighborhoods. *Gizmodo* (blog), 12 February. https://gizmodo.com/how-we-determined-predictive-policing-software-dispropo-1848139456

Metcalf, Jacob, Emanuel Moss, Elizabeth Anne Watkins, Ranjit Singh, and Madeleine Clare Elish (2021). Algorithmic impact assessments and accountability: The co-construction of impacts. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*. FAccT '21. New York: Association for Computing Machinery, pp. 735–46. https://doi.org/10.1145/3442188.3445935

Metz, Thaddeus (2021). African reasons why AI should not maximize utility. In *African Values, Ethics, and Technology: Questions, Issues, and Approaches*, Beatrice Dedaa Okyere-Manu, ed., pp. 55–72. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-70550-3_4

Middleton, Stuart E., Emmanuel Letouzé, Ali Hossaini, and Adriane Chapman (2022). Trust, regulation, and human-in-the-loop AI: Within the European region. *Communications of the ACM* 65, No. 4 (April), pp. 64–68. https://doi.org/10.1145/3511597

Milano, Silvia, Brent Mittelstadt, Sandra Wachter, and Christopher Russell (2021). Epistemic fragmentation poses a threat to the governance of online targeting. *Nature Machine Intelligence*, vol. 3, No. 6 (June), pp. 466–72. https://doi.org/10.1038/s42256-021-00358-3

Monge, Fernando (2022a). Cities have figured out a way to access private data in real time. *Datapoli's Newsletter* (blog), 25 March. https://fermonge.substack.com/p/cities-have-figured-out-a-way-to

Monge, Fernando (2022b). A new data deal: The case of Barcelona. *Datapoli's Newsletter* (blog), 25 February. https://fermonge.substack.com/p/a-new-data-deal-the-case-of-barcelona

Moreschi, Bruno, Gabriel Pereira, and Fabio G. Cozman (2020). The Brazilian workers in Amazon Mechanical Turk: Dreams and realities of ghost workers. *Contracampo – Brazilian Journal of Communication*, vol. 39, No. 1. http://dx.doi.org/10.22409/contracampo.v39i1.38252

Muhammad, Khan, Jaime Lloret, and Sung Wook Baik (2019). Intelligent and energy-efficient data prioritization in green smart cities: Current challenges and future directions. *IEEE Communications Magazine*, vol. 57, No. 2 (February), pp. 60–65. https://doi.org/10.1109/MCOM.2018.1800371

Murphy, Kevin P. (2012). *Machine Learning: A Probabilistic Perspective*. Cambridge: MIT Press. http://site.ebrary.com/id/10597102

Nagitta, Pross Oluka, Godfrey Mugurusi, Peter Adoko Obicci, and Emmanuel Awuor (2022). Human-centered artificial intelligence for the public sector: The gate keeping role of the public procurement professional. *Procedia Computer Science*, No. 200 (1 January), pp. 1084–92. https://doi.org/10.1016/j.procs.2022.01.308

Nahavandi, Darius, Roohallah Alizadehsani, Abbas Khosravi, and U. Rajendra Acharya (2022). Application of artificial intelligence in wearable devices: Opportunities and challenges. *Computer Methods and Programs in Biomedicine*, No. 213 (1 January), p. 106541. https://doi.org/10.1016/j.cmpb.2021.106541

Nam, Daisik, Hyunmyung Kim, Jaewoo Cho, and R. Jayakrishnan (2017). A model based on deep learning for predicting travel mode choice.*Semantics Scholar*. https://www.semanticscholar.org/paper/A-Model-Based-on-Deep-Learning-for-Predicting-Mode-Nam-Kim/56a8b850dcfa7900c59c79c3fa3e962c4aa5e916

Newaz, A. K. M. Iqtidar, Nur Imtiazul Haque, Amit Kumar Sikder, Mohammad Ashiqur Rahman, and A. Selcuk Uluagac (2020). Adversarial attacks to machine learning-based smart healthcare systems. *arXiv:2010.03671 [cs]* (7 October). http://arxiv.org/abs/2010.03671

Nguyen, Van Nhan, Robert Jenssen, and Davide Roverso (2018). Automatic autonomous vision- based power line inspection: A review of current status and the potential role of deep learning. *International Journal of Electrical Power & Energy Systems*, No. 99 (July), pp. 107–20. https://doi.org/10.1016/j.ijepes.2017.12.016

Nickerson, Raymond S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology*, vol. 2, No. 2, pp. 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nokia (2021). Nokia scene analytics. Nokia Digital Innovation Cloud (December). https://www.dac.nokia.com/applications/scene-analytics/

Noursalehi, Peyman, Haris N. Koutsopoulos, and Jinhua Zhao (2018). Real time transit demand prediction capturing station interactions and impact of special events. *Transportation Research Part C: Emerging Technologies* 97 (December), pp. 277–300. https://doi.org/10.1016/j.trc.2018.10.023

Nutkiewicz, Alex, Zheng Yang, and Rishee K. Jain (2018). Data-driven urban energy simulation (DUE-S): A framework for integrating engineering simulation and machine learning methods in a multi-scale urban energy modeling workflow. *Applied Energy*, No. 225 (September), pp. 1176–89. https://doi.org/10.1016/j.apenergy.2018.05.023

REFERENCES

Oard, Douglas, and Jinmook Kim (1998). Implicit feedback for recommender systems. AAAI Workshop on Recommender Systems, Madison, WI, pp. 81–83. https://scholarcommons.sc.edu/libsci_facpub/111/

OECD (Organisation for Economic Cooperation and Development) (2018). OECD regulatory policy outlook 2018. 10 October. https://doi.org/10.1787/9789264303072-en

OECD (2019). OECD AI principles overview (May). https://oecd.ai/en/ai-principles

OECD (2022). OECD Framework for the classification of AI systems. OECD Digital Economy Papers (22 February). https://doi.org/10.1787/cb6d9eca-en

OECD and United Nations Economic and Social Commission for Western Asia (UN ESCWA) (2021). Open government: Concept, definitions and implementation. In *The Economic and Social Impact of Open Government: Policy Recommendations for the Arab Countries*. https://www.oecd-ilibrary.org/governance/the-economic-and-social-impact-of-open-government_6b3e2469-en

Office of the Privacy Commissioner of Canada (OPC) (2021). News release: Clearview AI's unlawful practices represented mass surveillance of Canadians, commissioners say, 3 February. https://www.priv.gc.ca/en/opc-news/news-and-announcements/2021/nr-c_210203/

Omrani, Hichem (2015). Predicting travel mode of individuals by machine learning. *Transportation Research Procedia*, No. 10, pp. 840–49. https://doi.org/10.1016/j.trpro.2015.09.037

Open Mobility Foundation (2021). Recent court ruling supports mobility data specification. 26 February. https://www.openmobilityfoundation.org/an-important-development-in-los-angeles-for-the-mobility-data-specification-and-data-privacy/

Otieno, Fred, Nathan Williams, and Patrick McSharry (2018). Forecasting energy demand for microgrids over multiple horizons. In *2018 IEEE PES/IAS PowerAfrica*, Cape Town: IEEE, pp. 457–62. https://doi.org/10.1109/PowerAfrica.2018.8521063

Partnership on AI (2021). Fairer algorithmic decision-making and its consequences: Interrogating the risks and benefits of demographic data collection, use, and non-use. 2 December. https://partnershiponai.org/paper/fairer-algorithmic-decision-making-and-its-consequences/

Perera, Kasun S., Zeyar Aung, and Wei Lee Woon (2014). Machine learning techniques for supporting renewable energy generation and integration: a survey. In *Proceedings of the Second International Conference on Data Analytics for Renewable Energy Integration*. DARE'14. Cham: Springer, pp. 81–96. https://doi.org/10.1007/978-3-319-13290-7_7

Pertl, Michael, Kai Heussen, Oliver Gehrke, and Michel Rezkalla (2016). Voltage estimation in active distribution grids using neural networks. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*. Boston: IEEE, pp. 1–5. https://doi.org/10.1109/PESGM.2016.7741758

Petropoulos, Georgios (2022). The dark side of artificial intelligence: Manipulation of human behaviour. *Bruegel*, 2 February. https://www.bruegel.org/2022/02/the-dark-side-of-artificial-intelligence-manipulation-of-human-behaviour/

Prince, Anya, and Daniel Schwarcz (2020). Proxy discrimination in the age of artificial intelligence and big data. *Iowa Law Review*, vol. 105, No. 3. https://ilr.law.uiowa.edu/print/volume-105-issue-3/proxy-discrimination-in-the-age-of-artificial-intelligence-and-big-data

Prins, Corien, Haroon Sheikh, Erik Schrijvers, Eline de Jong, Monique Steijns, and Mark Bovens (2021). Mission AI: The new system technology. Netherlands Scientific Council for Government Policy (WRR). https://www.wrr.nl/binaries/wrr/documenten/rapporten/2021/11/11/opgave-ai-de-nieuwe-systeemtechnologie/Summary+WRRreport_Mission+AI_The+New+System+Technology_R105.pdf

PwC (2018). Fourth Industrial Revolution for the Earth: Harnessing artificial intelligence for the Earth. January. https://www.pwc.com/gx/en/services/sustainability/publications/ai-for-the-earth.html

Quinn, Thomas P., Stephan Jacobs, Manisha Senadeera, Vuong Le, and Simon Coghlan (2022). The three ghosts of medical AI: Can the black-box present deliver? *Artificial Intelligence in Medicine*, No. 124 (February), p. 102158. https://doi.org/10.1016/j.artmed.2021.102158

Rahim Taleqani, Ali, Jill Hough, and Kendall E. Nygard (2019). Public opinion on dockless bike sharing: A machine learning approach. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, No. 4 (April), pp. 195–204. https://doi.org/10.1177/0361198119838982

Ramchurn, Sarvapali, Perukrishnen Vytelingum, Alex Rogers, and Nicholas R. Jennings (2012). Putting the "smarts" into the smart grid: A grand challenge for artificial intelligence. *Communications of the ACM* 55, No. 4, pp. 86–97.

Raval, Noopur (2019). Automating informality: On AI and labour in the Global South. Global Information Society Watch. https://giswatch.org/node/6202

Regue, Robert, and Will Recker (2014). Proactive vehicle routing with inferred demand to solve the bikesharing rebalancing problem. *Transportation Research Part E: Logistics and Transportation Review*, No. 72 (December), pp. 192–209. https://doi.org/10.1016/j.tre.2014.10.005

Reuter, Tina (2020). Smart city visions and human rights: Do they go together ? Carr Center Discussion Paper Series,(Spring). https://carrcenter.hks.harvard.edu/files/cchr/files/CCDP_006.pdf

Richardson, Jordan P., Cambray Smith, Susan Curtis, Sara Watson, Xuan Zhu, Barbara Barry, and Richard R. Sharp (2021). Patient apprehensions about the use of artificial intelligence in healthcare. *npj Digital Medicine*, vol. 4, No. 1 (21 September), p. 140. https://doi.org/10.1038/s41746-021-00509-1

Rissland, Edwina L., Kevin D. Ashley, and R. P. Loui (2003). AI and law: A fruitful synergy. *Artificial Intelligence*, vol. 150, No. 1–2 (November), pp. 1–15. https://doi.org/10.1016/S0004-3702(03)00122-X

REFERENCES

Robinson, Caleb, Bistra Dilkina, Jeffrey Hubbs, Wenwen Zhang, Subhrajit Guhathakurta, Marilyn A. Brown, and Ram M. Pendyala (2017). Machine learning approaches for estimating commercial building energy consumption. *Applied Energy*, No. 208 (December), pp. 889–904. https://doi.org/10.1016/j.apenergy.2017.09.060

Rodríguez, Andrea G. (2021). AI ethics in olicy and action: City governance of algorithmic decision systems. COVID Briefs: Building back better: Post-pandemic city governance. CIDOB (Barcelona Centre for International Affairs), 15 December. https://dossiers.cidob.org/cities-in-times-of-pandemics/assets/pdf/Ai_ethics_in_policy_and_action.pdf

Rohaidi, Nurfilzah (2017). How the Maldives uses drones to fight climate change. November. *GovInsider* (3 November). https://govinsider.asia/inclusive-gov/undp-maldives-drones-climate-change-disaster-risk-maps/

Rosenblat, Alex and Luke Stark (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International Journal of Communication*, vol. 10, No. 27 (30 July). http://dx.doi.org/10.2139/ssrn.2686227

Rosenzweig, Cynthia, Joshua Elliott, Delphine Deryng, Alex C. Ruane, Christoph Müller, Almut Arneth, Kenneth J. Boote, et al. (2014). Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison. *Proceedings of the National Academy of Sciences of the United States of America*, vol. 111, No. 9 (4 March), pp. 3268–73. https://doi.org/10.1073/pnas.1222463110

Rothrock, Stephen M. (2021). Coronavirus, chip boom, and supply shortage: The new normal for global semiconductor manufacturing. *International Symposium on Microelectronics 2021*, No. 1 (1 October), pp. 000026–000030. https://doi.org/10.4071/1085-8024-2021.1.000026

Rovatsos, Michael, Brent Mittelstadt, and Ansgar Koene (2019). Landscape summary: Bias in algorithmic decision-making: What is bias in algorithmic decision-making, how can we identify it, and how can we mitigate it? UK Government (19 July). https://www.research.ed.ac.uk/en/publications/landscape-summary-bias-in-algorithmic-decision-making-what-is-bia

Russel, Stuart J, and Peter Norvig (2010). *Artificial Intelligence: A Modern Approach*. Upper Saddle River, NJ: Prentice Hall.

Safir, Inbal Naveh (2019). Using artificial intelligence as a tool for your local government. ICMA blog, 24 April. https://icma.org/blog-posts/using-artificial-intelligence-tool-your-local-government

Sagiroglu, Seref, and Duygu Sinanc (2013). Big data: A review. In *2013 International Conference on Collaboration Technologies and Systems (CTS)*, pp. 42–47. https://doi.org/10.1109/CTS.2013.6567202

Sahin, Kaan (2020). The West, China, and AI surveillance. *Atlantic Council* (blog), 18 December. https://www.atlanticcouncil.org/blogs/geotech-cues/the-west-china-and-ai-surveillance/

Saikia, Prarthana (2021). The importance of data drift detection that data scientists do not know. *Analytics Vidhya* (blog), 15 October. https://www.analyticsvidhya.com/blog/2021/10/mlops-and-the-importance-of-data-drift-detection/

Salamone, Francesco, Ludovico Danza, Italo Meroni, and Matteo Ghellere (2017). How to define the urban comfort in the era of smart cities through the use of the do-it-yourself approach and new pervasive technologies. Presented at the 4th International Electronic Conference on Sensors and Applications (ECSA 2017), 15–30 November. https://www.researchgate.net/publication/321076513_How_to_Define_the_Urban_Comfort_in_the_Era_of_Smart_Cities_through_the_Use_of_the_Do-It-Yourself_Approach_and_New_Pervasive_Technologies

Samimi, Amir, Kazuya Kawamura, and Abolfazl Mohammadian (2011). A behavioral analysis of freight mode choice decisions. *Transportation Planning and Technology*, vol. 34, No. 8 (December), pp. 857–69. https://doi.org/10.1080/03081060.2011.600092

Santos, Marcus A. G., Roberto Munoz, Rodrigo Olivares, Pedro P. Rebouças Filho, Javier Del Ser, and Victor Hugo C. de Albuquerque (2020). Online heart monitoring systems on the internet of health things environments: A survey, a reference model and an outlook. *Information Fusion*, No. 53 (1 January), pp. 222–39. https://doi.org/10.1016/j.inffus.2019.06.004

Santosh, K. C., Sameer Antani, D. S. Guru, and Nilanjan Dey, eds. (2019). *Medical Imaging: Artificial Intelligence, Image Recognition, and Machine Learning Techniques*. New York: CRC Press.

Schmitt, Lewin (2021). Mapping global AI governance: A nascent regime in a fragmented landscape. *AI and Ethics*, No. 2 (17 August), pp. 303–314. https://doi.org/10.1007/s43681-021-00083-y

Schwalbe, Nina, and Brian Wahl (2020). Artificial intelligence and the future of global health. *Lancet*, vol. 395, No. 10236, pp. 1579–86. https://doi.org/10.1016/S0140-6736(20)30226-9

Seo, Toru, Takahiko Kusakabe, Hiroto Gotoh, and Yasuo Asakura (2019). Interactive online machine learning approach for activity-travel survey. *Transportation Research Part B: Methodological*, vol. 123, No. C, pp. 362–73.

Shaban-Nejad, Arash, Martin Michalowski, and David L. Buckeridge (2018). Health intelligence: How artificial intelligence transforms population and personalized health. *npj Digital Medicine*, vol. 1, No. 1 (2 October), p. 53. https://doi.org/10.1038/s41746-018-0058-9

Shibasaki, Ryosuke, Satoru Hori, Shunji Kawamura, and Shigeyuki Tani (2020). Integrating urban data with urban services. In *Society 5.0: A People-Centric Super-Smart Society*, pp. 67–83. Singapore: Springer. https://doi.org/10.1007/978-981-15-2989-4_4

Shokri, Reza, Marco Stronati, Congzheng Song, and Vitaly Shmatikov (2016). Membership inference attacks against machine learning models. *arXiv:1610.05820 [cs.CR]* (18 October). https://doi.org/10.48550/ARXIV.1610.05820

Smith, Ember and Richard V. Reeves (2020). SAT math scores mirror and maintain racial inequity. *Brookings* (blog), 1 December. https://www.brookings.edu/blog/up-front/2020/12/01/sat-math-scores-mirror-and-maintain-racial-inequity/

Soleimanmeigouni, Iman, Alireza Ahmadi, and Uday Kumar (2018). Track geometry degradation and maintenance modelling: A review. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of Rail and Rapid Transit*, vol. 232, No. 1 (January), pp. 73–102. https://doi.org/10.1177/0954409716657849

REFERENCES

Southwest Research Institute (2016). SwRI developing methane leak detection system for DOE. Southwest Research Institute, 11 October. https://www.swri.org/press-release/swri-developing-methane-leak-detection-system-doe

Spatharou, Angela, Hieronimus Solveigh, and Jonathan Jenkins (2020). Transforming healthcare with AI: The impact on the workforce and organizations. *McKinsey*, 10 March. https://www.mckinsey.com/industries/healthcare-systems-and-services/our-insights/transforming-healthcare-with-ai

Sujith, A. V. L. N., Guna Sekhar Sajja, V. Mahalakshmi, Shibili Nuhmani, and B. Prasanalakshmi (2022). Systematic review of smart health monitoring using deep learning and artificial intelligence. *Neuroscience Informatics*, vol. 2, No. 3 (1 September). https://doi.org/10.1016/j.neuri.2021.100028

Sun, Li, and Fengqi You (2021). Machine learning and data-driven techniques for the control of smart power generation systems: An uncertainty handling perspective. *Engineering*, vol. 7, No. 9 (September), pp. 1239–47. https://doi.org/10.1016/j.eng.2021.04.020

Sun, Wenjuan, Paolo Bocchini, and Brian D. Davison (2020). Applications of artificial intelligence for disaster management. *Natural Hazards*, vol. 103, No. 3 (1 September), pp. 2631–89. https://doi.org/10.1007/s11069-020-04124-3

Sun, Yanshuo, Zhibin Jiang, Jinjing Gu, Min Zhou, Yeming Li, and Lei Zhang (2018). Analyzing high speed rail passengers' train choices based on new online booking data in China. *Transportation Research Part C: Emerging Technologies*, No. 97 (December), pp. 96–113. https://doi.org/10.1016/j.trc.2018.10.015

Suresh, Harini, and John V. Guttag (2021). A framework for understanding sources of harm throughout the machine learning life cycle. *EAAMO '21: Equity and Access in Algorithms, Mechanisms, and Optimization*, article no. 17, pp. 1–9. https://doi.org/10.1145/3465416.3483305

Taeihagh, Araz (2021). Governance of artificial intelligence. *Policy and Society*, vol. 40, No. 2, pp. 137–57. https://doi.org/10.1080/14494035.2021.1928377

Takahashi, Katsumi (2020). Social issues with digital twin computing. *NTT Technical Review*, vol. 18, No. 9 (September), pp. 36–39. https://www.ntt-review.jp/archive/ntttechnical.php?contents=ntr202009fa5.html

Tan, Jasper, Blake Mason, Hamid Javadi, and Richard G. Baraniuk (2022). Parameters or privacy: A provable tradeoff between overparameterization and membership inference. *arXiv:2202.01243 [stat.ML]* (2 February). https://arxiv.org/abs/2202.01243

Tanhaei, Bahareh, Ali Ayati, Manu Lahtinen, Behrooz Mahmoodzadeh Vaziri, and Mika Sillanpää (2016). A magnetic mesoporous chitosan based core-shells biopolymer for anionic dye adsorption: Kinetic and isothermal study and application of ANN. *Journal of Applied Polymer Science*, vol. 133, No. 22 (10 June). https://doi.org/10.1002/app.43466

Tavakoli, Reza, and Zeljko Pantic (2017). ANN-based algorithm for estimation and compensation of lateral misalignment in dynamic wireless power transfer systems for EV charging. In *2017 IEEE Energy Conversion Congress and Exposition (ECCE)*, Cincinnati, OH: IEEE, pp. 2602–9. https://doi.org/10.1109/ECCE.2017.8096493

Thakker, Dhavalkumar, Bhupesh Kumar Mishra, Amr Abdullatif, Suvodeep Mazumdar, and Sydney Simpson (2020). Explainable artificial intelligence for developing smart cities solutions. *Smart Cities*, vol. 3, No. 4 (December), pp. 1353–82. https://doi.org/10.3390/smartcities3040065

Trencher, Gregory, and Andrew Karvonen (2017). Stretching "smart": Advancing health and well- being through the smart city agenda. *Local Environment*, vol. 24, No. 7, pp. 610–27.

Tribby, Calvin P., Harvey J. Miller, Barbara B. Brown, Carol M. Werner, and Ken R. Smith (2017). Analyzing walking route choice through built environments using random forests and discrete choice techniques. *Environment and Planning B: Urban Analytics and City Science*, vol. 44, No. 6 (November), pp. 1145–67. https://doi.org/10.1177/0265813516659286

Trombetta, Paulo Henrique, Eriberto Nascimento, Elaine Carvalho da Paz, and Felipe do Valle (2018). Application of artificial neural networks for noise barrier optimization. *Environments*, vol. 5, No. 12 (10 December), p. 135. https://www.researchgate.net/publication/329543056_Application_of_Artificial_Neural_Networks_for_Noise_Barrier_Optimization

Tsapakis, Ioannis, and William H. Schneider (2015). Use of support vector machines to assign short-term counts to seasonal adjustment factor groups. *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2527, No. 1 (January), pp. 8–17. https://doi.org/10.3141/2527-02

Ugwudike, Pamela (2022). AI audits for assessing design logics and building ethical systems: The case of predictive policing algorithms. *AI and Ethics*, No. 2 (1 February), pp. 198–208. https://doi.org/10.1007/s43681-021-00117-5

UNITAC (n.d.). Using AI to map informal settlements in eThekwini, South Africa. United Nations Innovation Technology Accelerator for Cities. https://unitac.un.org/news/unitac-x-ethekwini-using-ai-map-informal-settlements-ethekwini-south-africa

United Nations (2017). New urban agenda. https://habitat3.org/the-new-urban-agenda/

United Nations (2019). Risk management in public administration in the context of the sustainable development goals. In *World Public Sector Report 2019*. https://www.un-ilibrary.org/content/books/9789210041409

United Nations, Committee on Economic Social and Cultural Rights (CESCR) (2000). General comment no. 14: The right to the highest attainable standard of health (art. 12 of the Covenant). https://www.refworld.org/docid/4538838d0.html

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2020). AI and the rule of law: Capacity building for judicial systems.4 November. https://en.unesco.org/artificial-intelligence/mooc-judges

UNESCO (2021). Recommendation on the ethics of artificial intelligence, 62201 §. https://unesdoc.unesco.org/ark:/48223/pf0000380455

REFERENCES

United Nations, UN-Habitat (2020). Centering people in smart cities: A playbook for local and regional governments. https://unhabitat.org/programme/people-centered-smart-cities/centering-people-in-smart-cities

United Nations, UN-Habitat (2021). A guide: Leveraging multi-level governance approaches to promote health equity.

University of Pretoria (2018). Artificial intelligence for Africa: An opportunity for growth, development, and democratisation. Access Partnership, 29 November.

Varshney, Kush R., and Homa Alemzadeh (2017). On the safety of machine learning: Cyber-physical systems, decision sciences, and data products. *arXiv:1610.01256 [cs.CY]* (22 August). https://doi.org/10.48550/arXiv.1610.01256

Vathoopan, Milan, Benjamin Brandenbourger, and Alois Zoitl (2016). A human in the loop corrective maintenance methodology using cross domain engineering data of mechatronic systems. In *2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1–4. https://doi.org/10.1109/ETFA.2016.7733603

Venugopalan, Subhashini, and Varun Rai (2015). Topic based classification and pattern identification in patents. *Technological Forecasting and Social Change*, No. 94 (May), pp. 236–50. https://doi.org/10.1016/j.techfore.2014.10.006

Victor, David G. (2019). How artificial intelligence will affect the future of energy and climate. *Brookings* (blog), 10 January. https://www.brookings.edu/research/how-artificial-intelligence-will-affect-the-future-of-energy-and-climate/

Voigt, Stefan, Thomas Kemper, Torsten Riedlinger, Ralph Kiefl, Klaas Scholte, and Harald Mehl (2007). Satellite image analysis for disaster and crisis-management support. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, No. 6 (June), pp. 1520–28. https://doi.org/10.1109/TGRS.2007.895830

Vosoughi, Soroush, Deb Roy, and Sinan Aral (2018). The spread of true and false news online. *Science*, vol. 359, No. 6380 (9 March), pp. 1146–51. https://doi.org/10.1126/science.aap9559

Voyant, Cyril, Gilles Notton, Soteris Kalogirou, Marie-Laure Nivet, Christophe Paoli, Fabrice Motte, and Alexis Fouilloy (2017). Machine learning methods for solar radiation forecasting: A review. *Renewable Energy*, No. 105 (May), pp. 569–82. https://doi.org/10.1016/j.renene.2016.12.095

Wan, Can, Jian Zhao, Yonghua Song, Zhao Xu, Jin Lin, and Zechun Hu (2015). Photovoltaic and solar power forecasting for smart grid energy management. *CSEE Journal of Power and Energy Systems*, vol. 1, No. 4 (December), pp. 38–46. https://doi.org/10.17775/CSEEJPES.2015.00046

Wan, Jiangwen, Yang Yu, Yinfeng Wu, Renjian Feng, and Ning Yu (2011). Hierarchical leak detection and localization method in natural gas pipeline monitoring sensor networks. *Sensors*, vol. 12, No. 1 (27 December), pp. 189–214. https://doi.org/10.3390/s120100189

Wang, Anna X., Caelin Tran, Nikhil Desai, David Lobell, and Stefano Ermon (2018). Deep transfer learning for crop yield prediction with remote sensing data. In *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*,. COMPASS '18. New York: Association for Computing Machinery, pp. 1–5. https://doi.org/10.1145/3209811.3212707

Wang, Shuangyuan, Ran Li, Adrian Evans, and Furong Li (2019). Electric vehicle load disaggregation based on limited activation matching pursuits. *Energy Procedia*, No. 158 (February), pp. 2611–16. https://doi.org/10.1016/j.egypro.2019.02.011

Wang, Zheye, Nina S. N. Lam, Nick Obradovich, and Xinyue Ye (2019). Are vulnerable communities digitally left behind in social responses to natural disasters? An evidence from Hurricane Sandy with Twitter data. *Applied Geography*, No. 108, pp. 1–8. https://doi.org/10.1016/j.apgeog.2019.05.001

Watts, Nick, W. Neil Adger, Sonja Ayeb-Karlsson, Yuqi Bai, Peter Byass, Diarmid Campbell-Lendrum, Tim Colbourn, et al. (2017). The Lancet countdown: Tracking progress on health and climate change. *The Lancet*, vol. 389, No. 10074 (March), pp. 1151–64. https://doi.org/10.1016/S0140-6736(16)32124-9

World Economic Forum (WEF) (2020). Global gender gap report 2020. http://reports.weforum.org/global-gender-gap-report-2020/dataexplorer

World Health Organization (WHO) (2021). Ethics and governance of artificial intelligence for health. https://www.who.int/publications-detail-redirect/9789240029200

World Health Organization (WHO) and UN-Habitat (2016). Global report on urban health: Equitable healthier cities for sustainable development.

Wirtz, Bernd W., Jan C. Weyerer, and Benjamin J. Sturm (2020). The dark sides of artificial intelligence: An integrated AI governance framework for public administration. *International Journal of Public Administration*, vol. 43, No. 9 (3 July), pp. 818–29. https://doi.org/10.1080/01900692.2020.1749851

Wong, Zoie S. Y., Jiaqi Zhou, and Qingpeng Zhang (2019). Artificial intelligence for infectious disease big data analytics. *Infection, Disease & Health*, vol. 24, No. 1 (1 February), pp. 44–48. https://doi.org/10.1016/j.idh.2018.10.002

Wu, Carole-Jean, Ramya Raghavendra, Udit Gupta, Bilge Acun, Newsha Ardalani, Kiwan Maeng, Gloria Chang, et al. (2022). Sustainable AI: Environmental implications, challenges and opportunities. *arXiv:2111.00364 [cs]* (9 January). http://arxiv.org/abs/2111.00364

Wu, Cathy, Aboudy Kreidieh, Kanaad Parvate, Eugene Vinitsky, and Alexandre M. Bayen (2022). Flow: A modular learning framework for mixed autonomy traffic. *IEEE Transactions on Robotics*, vol. 38, No. 2 (April), pp. 1270–86. https://doi.org/10.1109/TRO.2021.3087314

Wu, Cathy, Aboudy Kreidieh, Eugene Vinitsky, and Alexandre M. Bayen (2017). Emergent behaviors in mixed-autonomy traffic. In *Proceedings of the 1st Annual Conference on Robot Learning*, Proceedings of Machine Learning Research (PMLR), No. 78, pp. 398–407. https://proceedings.mlr.press/v78/wu17a.html

Wynants, Laure, Ben Van Calster, Gary S. Collins, Richard D. Riley, Georg Heinze, Ewoud Schuit, Marc M. J. Bonten, et al. (2020). Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. *BMJ*, No. 369 (7 April). https://doi.org/10.1136/bmj.m1328

REFERENCES

Xie, Yanhua, Anthea Weng, and Qihao Weng (2015). Population estimation of urban residential communities using remotely sensed morphologic data. *IEEE Geoscience and Remote Sensing Letters*, vol. 12, No. 5 (May), pp. 1111–15. https://doi.org/10.1109/LGRS.2014.2385597

Xu, Huan, and Shie Mannor (2012). Robustness and generalization. *Machine Learning*, vol. 86, No. 3 (1 March), pp. 391–423. https://doi.org/10.1007/s10994-011-5268-1

Xu, Zhenxing, Chang Su, Yunyu Xiao, and Fei Wang (2021). Artificial intelligence for COVID-19: Battling the pandemic with computational intelligence. *Intelligent Medicine*, vol. 2, No. 1 (21 October). https://doi.org/10.1016/j.imed.2021.09.001

Yang, Lei, Xin Liu, Weiqiang Zhu, Liang Zhao, and Gregory C. Beroza (2022). Toward improved urban earthquake monitoring through deep-learning-based noise suppression. *Science Advances*, vol. 8, No. 15 (13 April). https://www.science.org/doi/10.1126/sciadv.abl3564

Yigitcanlar, Tan, Juan M. Corchado, Rashid Mehmood, Rita Yi Man Li, Karen Mossberger, and Kevin Desouza (2021). Responsible urban innovation with local government artificial intelligence (AI): A conceptual framework and research agenda. *Journal of Open Innovation: Technology, Market, and Complexity*, No. 7. https://www.mdpi.com/2199-8531/7/1/71/pdf

You, Jiaxuan, Xiaocheng Li, Melvin Low, David Lobell, and Stefano Ermon (2017). Deep Gaussian process for crop yield prediction based on remote sensing data. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. AAAI'17. San Francisco: AAAI Press, pp. 4559–65.

Yu, Jiafan, Zhecheng Wang, Arun Majumdar, and Ram Rajagopal (2018). DeepSolar: A machine learning framework to efficiently construct a solar deployment database in the United States. *Joule*, vol. 2, No. 12 (December), pp. 2605–17. https://doi.org/10.1016/j.joule.2018.11.021

Zaki, Mohamed H., and Tarek Sayed (2016). Automated cyclist data collection under high density conditions. *IET Intelligent Transport Systems*, vol. 10, No. 5 (June), pp. 361–69. https://doi.org/10.1049/iet-its.2014.0257

Zeng, Daniel, Zhidong Cao, and Daniel B. Neill (2021). Chapter 22 – Artificial intelligence-enabled public health surveillance—from local detection to global epidemic monitoring and control. In *Artificial Intelligence in Medicine*, Lei Xing, Maryellen L. Giger, and James K. Min, eds. Academic Press, pp. 437–53. https://doi.org/10.1016/B978-0-12-821259-2.00022-3

Zhang, Wenwen, Caleb Robinson, Subhrajit Guhathakurta, Venu M. Garikapati, Bistra Dilkina, Marilyn A. Brown, and Ram M. Pendyala (2018). Estimating residential energy consumption in metropolitan areas: A microsimulation approach. *Energy*, No. 155 – (July), pp. 162–73. https://doi.org/10.1016/j.energy.2018.04.161

Zhang, Xiao, Gabriela Hug, J. Zico Kolter, and Iiro Harjunkoski (2016). Model predictive control of industrial loads and energy storage for demand response. In *2016 IEEE Power and Energy Society General Meeting (PESGM)*. Boston: IEEE, pp. 1–5. https://doi.org/10.1109/PESGM.2016.7741228

Zhang, Xinyi, Chengfang Fang, and Jie Shi (2021). Thief, beware of what get you there: Towards understanding model extraction attack. *arXiv:2104.05921 [cs]* (12 April). http://arxiv.org/abs/2104.05921

Zheng, Yongqing, Han Yu, Lizhen Cui, Chunyan Miao, Cyril Leung, and Qiang Yang (2018). SmartHS: An AI platform for improving government service provision. *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, No. 1 (27 April). https://doi.org/10.1609/aaai.v32i1.11382

Zhou, Li, and Margarita Sordo (2021). Chapter 5 – Expert systems in medicine. In *Artificial Intelligence in Medicine*, Lei Xing, Maryellen L. Giger, and James K. Min, eds. Academic Press, pp. 75–100. https://doi.org/10.1016/B978-0-12-821259-2.00005-3

Zissis, Georges (2019). The R3 concept: Reliability, robustness, and resilience [President's Message]. *IEEE Industry Applications Magazine*, vol. 25, No. 4, pp. 5–6. https://doi.org/10.1109/MIAS.2019.2909374

Zytek, Alexandra, Dongyu Liu, Rhema Vaithianathan, and Kalyan Veeramachaneni (2021). Understanding the usability challenges of machine learning in high-stakes decision making. *arXiv:2103.02071 [cs.HC]* (2 March). https://arxiv.org/abs/2103.02071

Mila

UN-HABITAT
FOR A BETTER URBAN FUTURE