

Lane, Lottie

## Article

# Preventing long-term risks to human rights in smart cities: A critical review of responsibilities for private AI developers

Internet Policy Review

## Provided in Cooperation with:

Alexander von Humboldt Institute for Internet and Society (HIIG), Berlin

*Suggested Citation:* Lane, Lottie (2023) : Preventing long-term risks to human rights in smart cities: A critical review of responsibilities for private AI developers, Internet Policy Review, ISSN 2197-6775, Alexander von Humboldt Institute for Internet and Society, Berlin, Vol. 12, Iss. 1, pp. 1-30, <https://doi.org/10.14763/2023.1.1697>

This Version is available at:

<https://hdl.handle.net/10419/271328>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/3.0/de/legalcode>



# Preventing long-term risks to human rights in smart cities: A critical review of responsibilities for private AI developers

**Lottie Lane** *University of Groningen* [c.l.lane@rug.nl](mailto:c.l.lane@rug.nl)

**DOI:** <https://doi.org/10.14763/2023.1.1697>

**Published:** 31 March 2023

**Received:** 16 September 2022 **Accepted:** 23 November 2022

**Funding:** This research was funded by the Dutch Sectorplan for Social Sciences and Humanities.

**Competing Interests:** The author has declared that no competing interests exist that have influenced the text.

**Licence:** This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 License (Germany) which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. <https://creativecommons.org/licenses/by/3.0/de/deed.en>  
Copyright remains with the author(s).

**Citation:** Lane, L. (2023). Preventing long-term risks to human rights in smart cities: A critical review of responsibilities for private AI developers. *Internet Policy Review*, 12(1). <https://doi.org/10.14763/2023.1.1697>

**Keywords:** Smart cities, Artificial intelligence, Human rights, AI Act, Corporate sustainability

**Abstract:** Privately developed artificial intelligence (AI) systems are frequently used in smart city technologies. The negative effects of such systems on individuals' human rights are increasingly clear, but we still only have a snapshot of their long-term risks to human rights. The central role of AI businesses in smart cities places them in a key position to identify, prevent and mitigate risks posed by smart city AI systems. The question arises as to how such preventive responsibilities are articulated in international and European governance initiatives on AI and corporate responsibility, respectively. This paper addresses the questions regarding: (1) the Organization for Economic Cooperation and Development's 'Business and Finance Outlook 2021: AI in Business and Finance'; (2) the EU's proposed 'AI Act'; and (3) the EU's 'Proposal for a Directive on corporate sustainability due diligence'. The paper first discusses the role of private AI developers in smart cities and the relevant limitations of applicable legal frameworks (section 1). Section 2 categorises long-term risks to human rights posed by the private development of smart city AI systems. Section 3 discusses how preventive responsibilities in the three initiatives reflect considerations of long-term risks. Critical observations and recommendations are provided in section 4, and conclusions are in section 5.

This paper is part of **Future-proofing the city: A human rights-based approach to governing algorithmic, biometric and smart city technologies**, a special issue of *Internet Policy Review* guest-edited by Alina Wernick and Anna Artyushina.

## Introduction

Artificial intelligence<sup>1</sup> (AI) systems are increasingly used in smart city contexts<sup>2</sup> (Herath & Mittal, 2022). The fast-paced development and deployment of smart city AI systems for new use cases raises significant questions concerning long-term risks posed to human rights and the responsibilities of the many private sector actors developing such systems. Corporate responsibility for human rights requires, *inter alia*, that private developers conduct human rights due diligence to “identify, prevent, mitigate and account for how they address their adverse human rights impacts”, (HRDD; Ruggie, 2011, Principle 17; section *Limitations of legal frameworks* below). In the context of smart city AI, HRDD should be future-proof and “resilient to a range of different plausible scenarios and that consider potential cascading impacts” (Allison-Hope & Hodge, 2018, p. 7). This requires paying explicit attention to long-term risks to human rights in applicable standards.

This paper is a response to the call for “‘anticipatory’ approaches to the study of responsible AI” (Whittlestone & Avin, n.d.). It bridges discourses on smart cities, AI and business and human rights, addressing the specific issue of corporate responsibility, which is often overlooked as a preventive approach in discussions about (EU) law and technology. Indeed, explicit discussion of “corporate responsibility” in the prevention of human rights interference caused by AI are relatively scarce, including in many AI governance initiatives.<sup>3</sup> Instead, although some may articulate human rights-related standards for private developers of AI, ex ante/preventive approaches to AI risk-management are usually discussed in terms of ethical principles (e.g. transparency and accountability), particular areas of law (e.g. privacy or data protection) or specific measures to be followed (e.g. ethical impact assess-

1. This paper adopts the definition of the UN Education, Scientific and Cultural Organization (2021) whereby AI systems are “technological systems which have the capacity to process information in a way that resembles intelligent behaviour, and typically includes aspects of reasoning, learning, perception, prediction, planning or control”.
2. There is no universally agreed definition of “smart cities” (e.g. Batty, 2020; Hollands, 2008, p. 306; Kitchen, 2015). This paper follows the European Commission’s (n.d.) definition referring to “place[s] where traditional networks and services are made more efficient with the use of digital solutions for the benefit of its inhabitants and business”.
3. A study of 111 AI governance initiatives conducted in July 2022 found that only 10 mentioned “corporate responsibility” explicitly, of which only two were European initiatives (Lane, 2023).

ments, human oversight and auditing) (Lane, 2023).

The present paper fills this gap, reviewing risk-management provisions in: (1) the Organization for Economic Cooperation and Development's (OECD) 'Business and Finance Outlook 2021: AI in Business and Finance' (2021); (2) the EU's proposed 'Artificial Intelligence Act' (AIA; 2022); and (3) the EU's 'Proposal for a Directive on corporate sustainability due diligence' (CSDDD; 2022). These initiatives have not yet been critically analysed from the perspective of corporate responsibility (the AIA), the development of smart city AI systems (the CSDDD) and/or in light of long-term risks (all three initiatives).

The risks that AI poses to human rights are well documented (Donahoe & Metzger, 2019; European Economic and Social Committee, 2016; Human Rights Council, 2021). In the context of smart cities, literature tends to focus on risks to privacy and data protection (Eckhoff & Wagner, 2018; Edwards, 2016; Kitchin et al., 2018; Voorwinden, 2021). Although the findings can be applied to all human rights, this paper pays more attention to non-discrimination in the provision of essential public services. Smart cities are often introduced to provide services "such as education, healthcare, sanitation, drinking water, and mobility", in an equitable manner (Jiang et al., 2022, p. 1639). However, many smart cities have failed to live up to expectations (Sengupta & Sengupta, 2022) and significant risks are posed to human rights closely related to these services (Marlies Hesselman et al., 2017).

Discrimination arising from smart city AI in the provision of public services could concern:

- Education, e.g. the allocation of pupils and resources to schools;
- Healthcare, e.g. the placing of individuals on transplant lists (Babic et al., 2020), the distribution of medical supplies/services;
- The allocation and supervision of welfare benefits within a municipality, as reflected in the Dutch 'SyRI' case (*NJCM et al. v The Dutch State (SyRI)*, 2020; Philip Alston, 2019; Rachovitsa & Johann, 2022) and the Dutch childcare benefits scandal (Amnesty International, 2021; Björn ten Seldam & Alex Brenninkmeijer, 2021; Henley, 2021).

Even seemingly "mundane" impacts of AI, such as more diligently reporting potholes in roads situated in more affluent or digitally advanced neighbourhoods, can exacerbate existing societal inequalities and could lead to discrimination against certain groups (Pellegrin et al., 2021, p. 35). Discrimination may not always be immediately evident. For example, although some data (e.g. regarding individuals' ethnicity) may not be collected for a smart city AI system, an algorithm may "infer

such characteristics from proxies (country of birth of a person or their parents, postcode, search interests etc)”, which could still result in indirect discrimination that is amplified when used at scale (Christophe Lacroix, 2020, para. 68; Cobbe et al., 2020).

This article focuses on the preventive responsibilities/obligations<sup>4</sup> (McCorquodale & Nolan, 2021) of private businesses developing smart city AI systems to take ex ante measures to prevent and mitigate the realisation of risks to human rights caused by their product when it is (mis)used by their (public) customers, intentionally or not.

Section 1 discusses the role of private AI developers in smart cities and correlating limitations of existing legal frameworks. Section 2 categorises long-term risks in this context as: (1) uncertainties in developments of AI technologies; (2) the unpredictability of certain AI systems; and (3) unforeseen uses of AI systems. Section 3 provides a doctrinal analysis of corporate responsibilities/obligations for private developers of smart city AI systems in the abovementioned initiatives. Section 4 provides recommendations for law- and policy-makers, as well as for businesses regarding standards for the prevention of long-term risks to human rights. These include explicitly referring to long-term risks, employing leverage with business relationships and considerations for human rights impact assessments. Conclusions are drawn in section 5.

## **Section 1: Private businesses, smart city AI systems and human rights law**

### **Private AI businesses and smart cities**

Public reliance on privately developed smart city AI solutions has progressed significantly (Goodman, 2020, p. 381) due to financial austerity, the promise of heightened efficiency through AI solutions, the vast technical expertise of the private sector and the shrinking capacity of the public sector to tackle socially and technically complex urban issues (Kitchin et al., 2017, p. 3, 2018, p. 6).

The role of private businesses in smart cities is context-specific (Jiang et al., 2022). Private businesses may be contracted by cities to use data analytics to allocate public resources, including welfare services (Goodman, 2020, p. 824). They may

4. The terms responsibilities and obligations are distinguished in this paper to reflect the fact that under international and European human rights law, unlike legally binding obligations, responsibilities are considered to impose non-binding standards on their addressees. Both types of standards are included in the paper's core analysis in section 3.

‘equip and orient emergency services and the police force’ or even develop new public services (Voorwinden, 2021, p. 442). Privately developed AI can change how public mandates are fulfilled (Ranchordás & Goanta, 2020) and in some situations the public sector is rendered a mere agent of corporations that take over the production and provision of public goods (Kempin Reuter, 2020, p. 4). This is sometimes true even in the governance of smart cities (Hollands, 2015; Jiang et al., 2022).

Economic interests are often the driving force behind privately developed smart city AI, which can favour deregulation, privatisation and more open economies, weakening oversight and enabling more efficient capital accumulation (Kitchin, 2015, p. 132) to the detriment of human rights considerations (Hollands, 2015). The human rights risks in privatising public services has long been a concern (Hal-lo de Wolf, 2012; Marlies Hesselman et al., 2017; UN Committee on Economic, Social and Cultural Rights, 1999, 2017). The smart city context raises new questions of what standards should be placed on private AI developers to prevent and mitigate ensuing human rights violations given their control over the design and workings of smart city AI systems.

## Limitations of legal frameworks

Additionally, the human rights responsibilities of businesses developing smart city AI are somewhat uncertain (Lane, 2022, 2023; Maas, 2019). A lack of binding sources on AI under international human rights law necessitates authoritative interpretations to explain how more general human rights obligations apply in the context of AI. Such interpretations are relatively sparse, focus on the conduct of States much more than businesses and do not address smart cities specifically, although several do address use cases found in smart cities and/or rights at stake in this context (Human Rights Council, 2021; UN Committee on the Rights of the Child, 2021).

Further, businesses do not currently have binding human rights obligations at the international level. Ongoing legal initiatives geared towards better corporate respect and accountability for human rights include a draft binding international treaty on business and human rights (Open-Ended Intergovernmental Working Group on Transnational Corporations and other Business Enterprises with respect to Human Rights, 2021) and the proposed EU directive on corporate sustainability due diligence (Council of the European Union, 2022; section *Preventive corporate responsibilities in the CSDDD* below). However, we still largely rely on non-binding standards on business and human rights, the most authoritative of which are the

United Nations' Guiding Principles on Business and Human Rights (UNGPs; McCorquodale & Nolan, 2021; Ruggie, 2011). The Principles lay down the essence of corporate responsibility for human rights, with HRDD at their core. According to Principle 17, HRDD requires businesses to “asses[s] actual and potential human rights impacts, integrat[e] and ac[t] upon the findings, trac[k] responses, and communicat[e] how impacts are addressed”. The UNGPs apply to all business enterprises regardless of sector. They therefore remain general, and guidance is necessary to explain their application to different contexts, including smart cities and AI (Lane, 2022). This task has to some extent been taken up by organisations such as BSR (Allison-Hope & Hodge, 2018) and the OECD (2021; section *Preventive corporate responsibility in the OECD's guidance* below).

## Section 2: Categorisation of long-term risks to human rights

Several long-term risks to human rights posed by AI have been identified in literature. This includes the unforeseen exploitation of personal data stored for a long time (Human Rights Council, 2021, para. 14), secrecy amongst private AI actors potentially leading to increased lack of control over time, “increased risk of government and corporate surveillance, and increased risks of security breaches” (Qarri & Gill, 2022, p. 12).

Although AI is not itself a new phenomenon, its development and deployment in a wide variety of new use cases within the smart city context has taken place over a relatively short period of time. Additionally, smart cities are characterised by complex networks of actors interacting with one another with changing roles and capacities. This provides relatively little insight into the longer-term risks to human rights posed by such AI systems, which require further investigation. Unlike more immediate and short-term risks, which may be more quickly and easily identified, long-term risks may only become apparent after significant study is undertaken to understand risks and their management (All-Party Parliamentary Group on Artificial Intelligence, 2019), after thorough analysis of unintended consequences of a system or once the system has already been deployed (Qarri & Gill, 2022). The “possible unknown unknowns and ‘black swans’” concerning long-term and future risks must be acknowledged (Muller, 2020).

Three key challenges to identifying long-term risks can be identified, which can also be considered long-term risks to human rights themselves: (1) uncertainties in developments of AI technologies (Allison-Hope & Hodge, 2018; Sharif & Pokharel, 2022); (2) the unpredictability of certain AI systems (Edwards & Veale, 2017); and



(3) unforeseen uses of AI systems (see also Hacker & Neyer, 2023 in this special issue). These three categories are briefly elaborated upon below. While safeguards can alleviate each challenge/risk to some degree, the risks render effective regulation of such systems more perplexing – the ‘abstract nature’ of many long-term risks and concerns regarding AI challenge our ability to understand how various systems should be developed, deployed and governed (Whittlestone & Avin, n.d.). The preventive human rights responsibilities of corporations may contribute in this respect.

## **Uncertainties in developments of AI technologies**

There is significant speculation about AI’s future development, which could raise new and as yet unknown risks to human rights. Although many commentators are sceptical that cognitive artificial general intelligence will become a reality and pose existential threats to human rights (Ahmad et al., 2021; Larson, 2021; Shanahan, 2015), there have been impressive developments in artificial narrow intelligence over the last years that were not envisaged in the earlier days of AI. This raises questions as to the future capabilities of AI systems and the speed at which they will be developed – throughout its history, this has fluctuated and has depended on developments in other technologies (Council of Europe, n.d.).

Further, as capabilities of AI systems continue to develop, what constitutes AI also develops (UN Educational, Scientific and Cultural Organization, 2021, para. 2). This poses difficulties to the effective, long-term regulation of AI, which must be defined in a future proof manner (see sections 3 and 4 below).

## **Unpredictability of certain AI systems**

The level and complexity of risks evolve alongside technology (Sharif & Pokharel, 2022) with some AI systems being infamously complex, opaque and difficult to predict (Kroll et al., 2016). This can make it difficult to identify their impact in both the short- and long-term, and can limit the ability, even of designers and developers of systems, to comprehend whether AI systems embody relevant values (Helbing et al., 2021; cited in Umbrello & van de Poel, 2021, p. 14; Yampolskiy, 2019). It is particularly worrying when such systems are used to make significant decisions affecting human rights, such as the distribution of public services and welfare benefits, as may be the case in smart cities (Ahmad et al., 2021). Lack of an accessible explanation of how a system works and/or came to a given output could hinder a victim’s ability to challenge reliance on that output (Roig, 2017), posing a significant obstacle to achieving accountability for harm and a victim’s right to an



effective remedy, should interference with human rights materialise. In the long-term, it is crucial to bear in mind that a lack of accountability for businesses developing smart city AI leaves space for them to continue building systems that could have a negative effect on human rights in the future.

## **Unforeseen uses of AI systems**

In addition, some AI systems designed for one purpose are ultimately used for another purpose unforeseen by the developers of the system. The notion of function creep (Hacker & Neyer, 2023 in this special issue) has evolved and is often discussed in the context of data protection (Koops, 2021). This is a central concern in relation to smart city AI systems, where data stored for internal use by one public authority may be shared with another authority. It may even be used to train an automated decision-making system for public authorities, taking risks beyond the sharing of personal data into the realm of rights associated with the new use case. Alternatively, a system designed and developed for use outside of the smart city context may be used in the provision of public services or law enforcement, raising new concerns as to its potential negative impact on related human rights such as housing, water, an adequate standard of living and discrimination.

## **Long-term risks in the smart city context**

Alongside risks concerning the nature and use of AI, the specific context of smart cities must be borne in mind. Rolling out a smart city AI system that causes discrimination on a city-wide basis over a number of years could have an amplifying effect over time, and could potentially lead to broader impacts than the immediate discriminatory effect. As Pellegrin et al. (2021) note, smart city AI could create “monopoly situations”, leaving some parts of a city’s population behind. One could imagine, for instance, that reliance on an AI system causing discrimination against people of a certain nationality or race in relation to their education could have an impact on their well-being in the long-term, for example, regarding future job opportunities and their standard of living.

Smart city AI systems addressing the complete lifecycle of a system, from conception/design to deployment and monitoring, is often emphasised as important (European Commission, 2020; High-Level Expert Group on Artificial Intelligence, 2019; OECD, 2019). To combat long-term risks, the stage of the lifecycle must not be considered in a vacuum – the actors involved at different stages should assess how the other stages in the lifecycle could, and have, unfold(ed). This could require, for instance, developers to contemplate the possible (mis-)uses of the tech-

nology, and users of systems to scrutinise what measures were taken to prevent and mitigate negative impacts during a system's development (e.g. whether a human rights impact assessment was conducted).

The ways in which long-term risks to human rights are reflected in the responsibilities/obligations in the three analysed initiatives are evaluated in the following paragraphs.

### Section 3: Preventive corporate responsibility in governance initiatives

A comprehensive analysis of the kaleidoscope of regulations and other governance initiatives adopted by a wide variety of actors and applicable to smart city AI (Goodman, 2020) is not possible within this paper. However, the OECD's guidance, the CSDDD and the AIA each contain preventive responsibilities/obligations applicable to private developers of AI and reflect aspects of corporate responsibility for human rights, even if implicitly. Table 1 demonstrates key comparative features of the initiatives that led to their selection for the analysis.

**TABLE 1:** Key comparative features of initiatives

	OECD BUSINESS AND FINANCE OUTLOOK	DIRECTIVE ON CORPORATE SUSTAINABILITY DUE DILIGENCE (CSDDD)	ARTIFICIAL INTELLIGENCE ACT (AIA)
Type of initiative	Non-binding, voluntary	Proposed EU Directive	Proposed EU Regulation
Geographical scope	International	EU, with potentially broader impacts	EU, with potentially broader impacts
Focus on AI	Yes	No	Yes

#### Preventive corporate responsibility in the OECD's guidance

Responsibilities of AI businesses are well explained in the OECD's 'AI in Business and Finance: Global Finance Outlook 2021'. The non-binding guidance builds on the OECD's previous work on AI (2019) and HRDD (2016). It addresses organisations and individuals developing, deploying or operating AI systems and can be

applied to all smart city AI systems.

The document aims to help the technology sector overcome the lack of international consensus on many human rights issues related to the development and use of AI. The OECD suggests HRDD activities for different actors in the supply chain, and at different stages in the AI lifecycle. It emphasises that AI HRDD is context-dependent, flexible and that businesses do not need to “disengage with high-risk activities”, but rather to tailor their context-dependent HRDD measures to the risks posed by a certain system, which they should prevent and mitigate (para. 3.3).

Some of the OECD’s standards could help to combat long-term risks despite not referring to them explicitly. First, HRDD should be ongoing. Following the UNGPs, the OECD emphasises that HRDD is not a tick-box exercise to be completed once. Rather, AI businesses should continue HRDD measures over time to take into account any changes in context or systems that could alter the risks posed.

Second, the OECD recommends the development of “explainable” AI systems. This could improve trustworthiness and understanding of the risks of complex “black box” smart city AI systems that use machine and deep learning (Ahmad et al., 2021; Lane, 2023; Phillips et al., 2020; Vilone & Longo, 2021). Regarding the long-term risks identified in section 2: unpredictability and un-explainability are not synonymous (Yampolskiy, 2019). Nonetheless, developing explainable smart city AI systems and providing users with an explanation of how the systems work may help the users to identify and mitigate risks caused by their output and could encourage them to more critically engage with the system’s output. As mentioned in section *Unpredictability of certain AI systems*, it could also help victims to challenge a system, although the burden to do this should not lie only with victims – developers of smart city AI and the public bodies using them should have an appropriate role in the process of ensuring justice (Cobbe et al., 2020). Attention should also be paid to the limitations of explainability. In some cases this is due to the unpredictability of the use of an AI system (Abedin, 2022; Silver et al., 2016), but critics have voiced doubts as to the true usefulness of explainability and a “right to an explanation” of how decisions are made using AI (Edwards & Veale, 2017; Gryz & Rojszczak, 2021, citing Rudin, 2019, p. 206), suggesting that additional measures such as certification frameworks are necessary to balance competing interests and rights.

Third, employing leverage could improve the identification and prevention/mitigation of long-term risks. Leverage is the ability of an enterprise “to affect change in the wrongful practices of the entity that causes the harm” (OECD, 2011, Commen-

tary on General Policies para. 19). Leverage is key to HRDD for long-term risks. For example, during the deployment of AI developers have “a unique, ongoing relationship” with clients that is not found in other sectors, due to the need for customer support and software updates.

The OECD suggests limiting licensing renewal for end-users of (smart city) AI systems, essentially placing conditions on the licencing renewal of systems to prevent unintended (mis)use of a system, or to require that users avoid using a system for purposes incompatible with certain values (i.e. human rights). Other suggestions include developing a “kill switch” for certain features of a system and maintaining the right to terminate users’ access to systems in case of intentional misuse causing human rights abuse. The examples acknowledge that risks to human rights may emerge despite measures taken in the earlier stages of an AI system’s lifecycle, or once it is out of the developer’s direct control. They could also help to curb some unforeseen risks to human rights caused by misuse of an AI system and avoid negative instances of function creep.

Another important aspect of leverage proposed by the OECD is for AI businesses to take measures to ensure that business relationships throughout their value chain conduct HRDD themselves and follow responsible business practices. Due to an accompanying requirement of communication between stakeholders, this could result in risks that are unforeseeable to one actor in the value chain becoming more visible when identified by another actor in the chain, and ultimately aid in the prevention of long-term risks throughout the supply chain. However, effective leverage relies on businesses’ influence over actors in their usually very complex supply chains, which may be spread across multiple countries (Lane, 2022; Scherer, 2016).

Fourth, the OECD suggests that AI businesses train users of their systems. This could be key to preventing the long-term risks for unintentional future misuse of systems caused by a lack of understanding of AI systems, its capabilities, limitations and intended uses. The more a user understands about a system, the easier it should be to use it – or refrain from using it – in a responsible manner without posing (additional) risks to human rights. This is particularly relevant in the smart city context where public-private partnerships result in many privately developed systems being used by public actors without expertise in AI technologies.

Finally, the OECD’s standard of consulting with affected stakeholders could help to identify long- and short-term risks to human rights. Such consultation is mentioned in the UNGPs as well as in many AI governance initiatives (Lane, 2023). Typically, reference is made to mitigating bias and/or the discriminatory effects of re-

liance on AI systems, which could be achieved by engaging affected stakeholders during the design and development stages of a system's lifecycle, *inter alia* by ensuring that the data used to train the system is representative. This is crucial in relation to AI systems used in the context of essential public services, as international human rights law requires that their provision be accessible without discrimination (e.g. UN Committee on Economic, Social and Cultural Rights, 1999).

## **Preventive corporate responsibilities in the CSDDD**

In February 2022, the European Commission submitted a proposal for the CSDDD in the form of an EU directive. In December 2022, the Council of the EU adopted its general approach, which forms the basis of the following analysis. The CSDDD is applicable to 'companies' meeting certain criteria regarding their size, profit margin and location (Article 2). The CSDDD is not applicable to all private developers of AI, but would be applicable to some BigTech players in European smart cities.<sup>5</sup>

The CSDDD refers to the UNGPs and the OECD Guidelines for Multinational Enterprises (2011) and explicitly aligns with the latter (Recitals 5-6, 16, respectively). It has nevertheless been criticised for disregarding key principles within these initiatives, such as the context-dependent nature of HRDD based on factors such as a business' size (International Federation for Human Rights, 2022, p. 2). Nonetheless, the CSDDD follows the OECD in not requiring businesses to "guarantee, in all circumstances, that adverse impacts will never occur or that they will be stopped" (Recital 15).

Articles 5, 6, 7 and 10 of the CSDDD are the most relevant in the context of preventive responsibilities, explaining key requirements of HRDD as defined in Article 4(1). Article 5(1) requires companies to have a HRDD policy, which should include a description of its long-term approach to HRDD, processes and measures put in place, as well as a code of conduct. The policy should be updated every 24 months, requiring businesses to reassess the rules and principles they - and their business partners - should follow (Article 5(2)). This could encourage businesses to consider HRDD measures, including risk assessments, from a long-term perspective.

Article 6(1) requires that businesses take measures to identify actual or potential human rights and environmental impacts in their own operations, those of their

5. The European Parliament (2022) proposed that the production of AI be considered a "high-impact sector", to which lower thresholds apply concerning the applicational scope of the CSDDD (Recitals 21-22; Article 2(1)), but this does not feature in the Council's general approach.

subsidiaries and of their business partners within their value chain. A heavily criticised earlier draft (European Commission, 2022) limited the responsibility to a business' "established relationships", defined by reference to their intensity or duration, as well the relation's position within the value chain (Article 3(f)). This risked limiting HRDD to relations only one or two tiers from the original company, significantly decreasing the impact of the CSDDD's requirements relating to leverage (discussed below) because it could encourage them to pursue short-term relationships to limit their HRDD obligations (Forest Peoples Programme, 2022). As a sector with notoriously complex and opaque supply chains comprising a wide variety of relationships (Lane, 2022; Organization for Economic Cooperation and Development, 2021; Scherer, 2016), strong HRDD that applies throughout an AI business' supply chain is vital to the successful prevention of (long-term) risks to human rights. Replacing the narrow concept of "business relationship" with the broader concept of "business partner", whether direct or indirect (Article 3(e)), could therefore impose a more effective obligation in the context of smart city AI.

Article 7 requires the adoption of appropriate measures to prevent or adequately mitigate adverse impacts. Article 7(2) specifies a number of measures that could contribute to protection from long-term risks. Notably, companies should seek 'contractual assurances' from direct business partners that they will comply with the business' code of conduct, and require the same of their partners throughout their own value chain (Article 7(2)(b)). Importantly, this "contractual cascading" must be accompanied by "appropriate measures to verify compliance" (Article 7(3)), placing a sort of enforcement requirement on businesses. The potential of this clause is limited by the fact that only "direct" business partners are covered, which are limited to those with a commercial agreement (Article 3(e)(i)). This may not be the case in relationships between businesses and public actors (International Federation for Human Rights, 2022, p. 3), potentially limiting its impact in the smart city context. Care must also be taken to ensure that contractual cascading does not enable companies to contractually confer responsibility for HRDD to their business partners and avoid their own responsibility (Simmons & Simmons, 2022).

Further, under Article 10 businesses must periodically assess their HRDD implementation to verify they are properly identifying risks and that preventive or corrective measures are implemented, as well as determining how effective those measures have been. As explained above, this is crucial in relation to long-term risks.

Finally, the draft allows the Commission to issue guidelines for specific sectors or adverse impacts in order to support companies in their implementation of the CS-

DDD (Article 13). This opens the possibility for AI and smart city-specific guidance, and could draw on the OCED's guidance to provide a greater degree of legal certainty and cohesion. Guidance could be provided, for example, regarding whether the development of explainable AI and training for users of smart city AI systems would be appropriate measures to prevent or mitigate risks to human rights under Article 7.

## **Preventive corporate responsibility in the proposed AIA**

In April 2021, the first proposal of the AIA in the form of an EU regulation was adopted by the European Commission (for discussion, see Ebers et al., 2021; Edwards, 2022; Floridi, 2021; Veale & Borgesius, 2021). The AIA builds on previous EU initiatives to regulate AI, including the High-Level Expert Group on Artificial Intelligence's, Ethics Guidelines for Trustworthy AI (2019) and the European Commission's White Paper on AI (2020). Like the AIA itself, both documents emphasise the need to ground European AI in fundamental rights. The Council of the EU's general approach on the AIA (2022a) forms the basis of the following analysis. Given its specific focus on AI, the AIA (along with the OECD's guidance) could be used to help interpret the more general risk management standards of the CSDDD for AI companies falling within its scope.

The AIA claims to take a “future proof” approach, at least in defining AI (para. 1.2; Recital 6). This includes systems developed through machine learning approaches and logic- and knowledge-based approaches (para 1.1; Recitals 6a and 6b; Article 3(1)). To accommodate new technological developments, Article 4 allows the European Commission to adopt and implement acts to “further specify and update techniques” under these approaches (para. 1.2; Recital 6). This could help temper the uncertain future of AI developments mentioned in section *Uncertainties in developments of AI technologies* above.

The draft AIA imposes several preventive requirements on providers of “high-risk” AI systems. Following the proposal's definition of “high-risk” (Article 6; Annex III) and “artificial intelligence system” (Article 3(1)), these requirements may be applicable to a number of smart city contexts, such as some systems used in education, employment, access to and enjoyment of essential public and private services, law enforcement and the management and operation of critical infrastructure (Annex III paras. 2, 4, 5, 6; Sawhney, 2022). Article 7 allows the list of high-risk systems in Annex III to be revised by means of delegated acts, subject to new conditions in Article 7(3) to ensure the protection of fundamental rights in case of deletions.



There is no reference to corporate responsibility or HRDD in the proposal, but “high risk” AI systems are “those that have a significant harmful impact on the health, safety and fundamental rights of persons” (Recitals 27, 32). This is mirrored in the close connection between human rights (e.g. the rights to education, work and social security) and the high-risk systems listed in Annex III. Viewing this alongside the requirements concerning high-risk systems, we see a reflection of HRDD standards typically found in corporate responsibility initiatives such as the UNGPs and the OECD Guidelines.

The connection is most evident in Article 9 on the establishment, implementation, documentation and maintenance of a risk management system. Similar to the OCED initiative and the CSDDD, we see something like a human rights impact assessment here, which should be a “continuous, iterative process” to be updated regularly and cover a high-risk system’s entire lifecycle (Recital 42; Article 9(2)). As stressed above, this is important for preventing/mitigating long-term risks – the more a provider conducts such assessments, the more experience they will gain, and hopefully, as the long-term risks of AI are more thoroughly researched, they will become better at identifying them. Importantly for the smart city context, risk management should encompass the system’s interaction with the environment in which it is used (Recital 42) and therefore, arguably, the longer-term risks specific to certain smart city use cases. This goes further than the CSDDD and OECD’s standards to potentially assess how the impacts of systems leading to discriminatory access to certain services and resources could be amplified over time with consequences for a broader range of rights.

Unfortunately, the scope of the obligation to identify and analyse risks is quite limited, applying only to “known and foreseeable risks most likely to occur [...] in view of the intended purpose” of the system (Article 9(2)(a)). The wording here may have a negative impact on the identification and assessment of long-term risks, as they may be more difficult to foresee and their likelihood may be more challenging to determine. The limitation to “foreseeable” risks does not necessarily exclude long-term risks if these can be considered foreseeable, but the phrase “in view of the intended purpose” could be interpreted to exclude foreseeable misuse of a system. Article 9(2)(c) requires providers to evaluate “other possibly arising” risks based on the post-market monitoring system of high-risk systems required under Article 61 of the proposal.<sup>6</sup> This could allow risks that become visible at a later stage in the AI lifecycle to be assessed, going some way to dealing with the un-

6. Providers must also immediately report serious incidents arising from the use of their system to relevant market surveillance authorities (Article 62).

foreseen output of machine learning systems (Recital 78), and would entail providers establish “sophisticated communication channels across supply lines” (Cankett and Liddy, 2022). However, it does not *per se* extend the scope of Article 9(2)(a).

Overall, Article 9 displays three core aspects of HRDD regarding preventive responsibilities as found in the UNGPs and the OECD guidance: assessment of risks; measures taken in response to risks (Article 9(5)-(6); and ongoing assessments.<sup>7</sup> Tracking the effectiveness of measures taken could fall under the remit of future risk assessments (therefore having a preventive effect) given that they should be ongoing, but this could be made more explicit, for instance, with an obligation for providers to establish an internal audit function (Schuett, 2022).

Further, risk-management has influenced other provisions of the AIA (Mahler, 2022). For instance, Article 10 includes obligations on data governance and management, which could prevent or mitigate risks of discrimination – the ‘examination of possible biases that are likely to affect health and safety of natural persons or lead to discrimination prohibited by Union law’ (Article 10(2)(f)) is arguably a risk assessment measure. Unfortunately, the limitation of discrimination prohibited by Union law could exclude disproportionately negative effects on people on the basis of characteristics not protected by EU law (Wachter, 2022), such as income and marital status. The obligation that data sets be representative could also go some way to mitigating discrimination, as could the provision that providers be able to process special categories of personal data “in order to ensure [...] bias monitoring, detection and correction” for high-risk systems (Article 10(3) and (5); Recital 44; European Union Agency for Fundamental Rights, 2022).

Like the OECD guidance, the AIA includes requirements with respect to explainability. Article 13(1) requires that high-risk systems be designed and developed to ensure “sufficiently transparent” operation of them, “with a view to [...] enabling users to understand and use the system appropriately”. Article 13(2) then requires that “relevant, accessible and comprehensible” instructions be provided to users of high-risk systems. This includes information about a system’s capabilities, and any limitations and known or foreseeable circumstances linked to its intended purpose that could pose risks to fundamental rights (Article 13(3)(b)). As Francesco Sovrano et al. (2022, p. 131) notes, this is a form of “user-empowering explainability” that helps to ensure a system is used correctly. In light of long-term risks, it may be

7. The AIA also contains obligations concerning responding to (Article 9(5)-(6)) and communicating risk-management activities (Articles 11, 12, 23; Recital 46), discussion of which falls outside the scope of the article.

more useful regarding unforeseen uses of AI rather than the unpredictability of an AI system.

Article 13 is strengthened by human oversight requirements in Article 14, which aim to prevent or mitigate risks to fundamental rights caused by the use or reasonably foreseeable misuse of a system. The purpose of these requirements is, among others, to enable a user to interpret its results and to decide, when appropriate, not to use or follow, to disregard, override or reverse its output (Article 14(4); Sovrano et al. 2022, pp. 131-132). Unlike the OECD's initiative, Article 14 does not go so far as to impose a training obligation on providers, but it could go some way to mitigating the risks posed by unpredictable (use of) high-risk smart city AI systems falling under the AIA's scope. Article 14(3) foresees two main ways of achieving effective oversight: by the provider building measures into a system and/or identifying measures to be implemented by the user. Read together with Recital 50, which states that technical solutions to prevent harmful/undesirable behaviour could include "mechanisms enabling the system to safely interrupt its operation", the former could go so far as to include a "kill switch" akin to that suggested in the OECD's initiative. Measures implemented by the user could include checks to prevent "automation bias" by "over-relying on the output" of the system (Article 14(4)(c); Green, 2022).

Interestingly, although the proposal does not mention leverage explicitly, Article 27 requires distributors to refrain from putting a system on the market they think does not meet the proposal's risk-management requirements, to "take the corrective actions necessary to bring it into conformity with those requirements, to withdraw [the system] or recall it", or - crucially - to ensure such action from providers, importers or operators. In addition, in January 2023 co-rapporteurs of the European Parliament proposed that users of high-risk systems conduct a fundamental rights impact assessment (Bertuzzi, 2023). At the time of this writing, how the obligations would interact with those of providers remains unclear. It could be accompanied by an obligation inverse to Article 27 for providers to obtain confirmation from users before a system is procured, ensuring that they will conform with impact assessment requirements themselves.

## **Section 4: Critical observations and recommendations**

Each initiative assessed articulates some preventive responsibilities/obligations relevant to long-term risks to human rights posed by privately developed smart city AI systems. Key here are ongoing human rights impact assessments, explainable AI, leverage with business relationships and training of users. However,

stronger standards are needed for the identification, prevention and mitigation of these risks posed by smart city AI systems.

## General critical observations

Several points of critique can be made, with accompanying recommendations for law/policy-makers and private developers of smart city AI systems. The first concerns the scope of risk management measures. Applicable standards should clarify that impact assessments cover long-term risks. To increase businesses' chances of success with such impact assessments, guidance should be offered regarding how to identify (long-term) risks. Among others, inspiration could be taken from the EU AI Alliance (2018) and particularly the work of BSR (Allison-Hope & Hodge, 2018), which suggests the use of "scenario planning" and "future wheels" to try to foresee plausible long-term risks posed by AI systems. For the more generally applicable CSDDD, Article 13 could be utilised to provide such guidance in relation to smart city AI systems. This builds on recommendations such as those of Federik Zuiderveen Borgesius (2018), Nuno Gomes de Andrade and Verena Kontschieder (2021) that regulators "provide specific and detailed guidance on how to implement an [impact assessment] process, and release it alongside the law".

Due attention must also be paid to specific standards of human rights applicable to the smart city context, covering the full range of civil, political, economic, social and cultural rights impacted by smart city technologies. For instance, international human rights law requires that the provision of essential public services must remain accessible, affordable, adequate and of good quality, even when provided by private actors (UN Committee on Economic, Social and Cultural Rights, 2017). This would extend to the digitalisation of such services relying on (for example) automated decision-making systems to determine their distribution.

Additionally, attention must be paid to how assessment of and response to risks should be prioritised given that none of the initiatives analysed require the results of HRDD to eradicate all risks. Crucially, although the OECD suggests that risks are prioritised according to their severity, "[a]ll impacts are expected to be addressed" (para. 3.3.3). This contrasts with the AIA, Article 9(2)(a) which only requires identification of (and therefore response to) risks "most likely to occur", although the notion of severity is reflected to some extent in the AIA's notion that high-risk systems concern "serious harm". In the CSDDD, both severity and likelihood are used as guideposts for prioritisation. In particular, businesses are expected to prioritise the most significant adverse impacts based on their gravity, the extent of harm (e.g. number of persons affected) and how difficult it is to essentially negate the

adverse impact (para. 17; Article 6a).

Following the UNGPs, the OECD's HRDD guidance and the CSDDD, an appropriate approach is based on the severity of risks, i.e. their scope, scale and irremediability (Ruggie, 2011, Commentary to Principle 14; Allison-Hope & Hodge, 2018, p. 7). Nevertheless, caution remains necessary when using the terminology of severity, as it is sometimes connected to the likelihood of a risk (e.g. Allison-Hope & Hodge, 2018, p. 7). It is important that corporate responsibility standards be clear and avoid fostering any preconceptions that the timeframe of risks necessarily affects the likelihood of their realisation.

## **Recommendations in the light of long-term risks**

Finally, when building smart city AI systems, private businesses need to take a holistic approach to risk management and HRDD. Breaking this down into more concrete suggestions according to the categorisation of long-term risks in section 2 above, law-makers and businesses should pay attention to the following:

### **Unforeseen use of AI systems**

Focus is placed, at least in the AIA, on foreseeable risks. It is both reasonable and practical that “unforeseeable risks” are not included in the requirements. Providers’ obligation under Article 14(2) to ensure that effective human oversight is possible during a system’s use extends to the “reasonably foreseeable misuse” of a system. However, beyond a brief explanation in Article 3(13), its meaning remains vague. For example, does reasonableness depend on a businesses’ size and/or resources to conduct an assessment into possible misuse of their system? Without providing some parameters of how reasonable foresight is measured, the AIA seems to allow providers to determine this for themselves. One way to clarify the matter would be to borrow from States’ due diligence obligations under international and European human rights law, a key aspect of which is reasonableness. What is reasonably to be expected of States is assessed alongside their knowledge/foresight of risks, their capacity, the interests at stake and the control of the State in a given situation (Malaihollo, 2021; Monnheim, 2021). While not directly transposable to AI businesses, these parameters could provide inspiration for defining what is a “reasonably foreseeable” risk or misuse of smart city AI.

Further, as explained above, the initiatives’ requirements related to leverage could go some way to addressing unpredictable and therefore unforeseeable (mis)uses of AI. To strengthen this further, we could learn from the GDPR, Article 5 of which provides that personal data can be collected only for specific purposes and may

only be further processed in a manner compatible with those purposes (Koops, 2021, p. 47). Smart city AI regulations, and (for example) contractual assurances with business relationships, could include a similar requirement, that a system may only be used for certain purposes compatible with certain values, including human rights.

Human rights impact assessment requirements for users could raise interesting questions concerning leverage. In order to address issues of leverage, developers could ask themselves the following, non-exhaustive questions: Who is the customer and how trustworthy are they? Are they known to have caused human rights abuse in the past? What influence does the developer have over them? Could this be increased? If not, should the business relationship and/or development of the AI system be terminated to safeguard human rights?<sup>8</sup>

Further, along the lines of future wheels and scenario planning (Allison-Hope & Hodge, 2018), what different scenarios could occur if certain decisions are made during different stages of the AI lifecycle? What other applications could the system have and in what ways could the system be misused? What could be the consequent impacts on human rights?

### **Unpredictability of AI systems**

It is important that developers consider how predictable a system is, and whether it is possible to make the system more explainable, bearing in mind the above mentioned caveats that come with this standard. In this respect, it is crucial that non-technical measures are also taken, for instance, requiring providers of systems to provide users of their systems with adequate information and training regarding these aspects of their systems, and to conduct effective post-market monitoring on an ongoing basis. Impact assessment requirements for users could also contribute to identifying risks and human rights impacts in a timelier manner.

### **Developments in AI technologies**

This category of long-term risks identified in section 2 receives the least attention in the initiatives analysed above. Nonetheless, it is important for developers to review how technological developments may be able to better prevent/mitigate negative human rights impacts, and whether new capabilities are able to offer more or less protection from human rights abuse. This is something that could be done during ongoing human rights impact assessments and the identification of appropriate risk management responses.

8. For further discussion see the UNGPs, Principle 19 and commentary thereto.

As mentioned above, it is crucial to remember that the recommendations here are not panaceas to the prevention and mitigation of risks to human rights. However, especially if taken together, they could contribute to the protection of human rights in the smart city context. Ultimately, given the lack of explicit discussion of long-term risks in the initiatives analysed, it is even more important to have effective accountability mechanisms in place to allow access to an effective remedy when risks materialise (e.g. UNGPs, Principle 22).

## Conclusion

The findings show that the three initiatives assessed articulate some corporate human rights responsibilities/obligations regarding long-term risks posed by smart city AI. Returning to the categorisation of risks in section 2, several conclusions can be reached. First, with respect to the uncertainty of developments in AI, the requirement in each initiative that HRDD is an ongoing process to be repeated over time is very helpful, as well as the “future proof” definition of AI in the AIA.

Second, the unpredictability of some systems can be combated by various requirements in the OECD and AIA, and could be interpreted into the CSDDD’s more general requirements. This concerns the explainability of systems and training of/providing information to users of systems, which could also contribute to minimising unforeseen (mis)uses of AI, especially when training on the intended use and limitations of a system is provided. Leverage is also key here, with particularly strong guidance from the OECD in this respect, a somewhat weaker version in the CSDDD and a very restricted form in the AIA. Returning to the example of non-discriminatory provision of public services, measures such as consultation with affected stakeholders and consideration of the amplifying effects of discrimination over time are also crucial.

Caution should nevertheless be taken to include more explicit recognition of long-term risks within applicable standards, to acknowledge the limits of existing and suggested measures and to avoid viewing the responsibilities/obligations of different actors in the AI supply chain/lifecycle in a vacuum. Finally, preventive responsibilities must be accompanied by effective corporate accountability standards, which should be the subject of further study.

---

## References

Abedin, B. (2022). Managing the tension between opposing effects of explainability of artificial



intelligence: A contingency theory perspective. *Internet Research*, 32(2), 425–453. <https://doi.org/10.1108/INTR-05-2020-0300>

Ahmad, K., Maabreh, M., Ghaly, M., Khan, K., Qadir, J., & Al-Fuqaha, A. (2021). *Developing future human-centered smart cities: Critical analysis of smart city security, interpretability, and ethical challenges* (arXiv:2012.09110). arXiv. <http://arxiv.org/abs/2012.09110>

Allison-Hope, D., & Hodge, M. (2018). *Artificial intelligence: A rights-based blueprint for business, paper 3: Implementing human rights due diligence* [Working Paper]. BSR. <https://www.bsr.org/reports/BSR-Artificial-Intelligence-A-Rights-Based-Blueprint-for-Business-Paper-03.pdf>

All-Party Parliamentary Group Artificial Intelligence & All-Party Parliamentary Group for Future Generations. (2019). *Long term artificial intelligence risks*. All-Party Parliamentary Group for Future Generations. <https://www.appgfuturegenerations.com/long-term-ai-risks>

Alston, P. (2019). *Brief by the United Nations Special Rapporteur on extreme poverty and human rights as Amicus Curiae in the case of NJCM c.s./De Staat der Nederlanden (SyRI) before the District Court of The Hague (Case number: C/09/550982/HA\_ZA\_18/388)*. United Nations Human Rights Office of the High Commissioner. <https://www.ohchr.org/sites/default/files/Documents/Issues/Poverty/Amicusfinalversionsigned.pdf>

Amnesty International. (2021). *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* [Report]. Amnesty International. <https://www.amnesty.org/en/documents/eur35/4686/2021/en/>

Babic, B., Cohen, I. G., Evgeniou, T., Gerke, S., & Trichakis, N. (2020, December 1). Can AI fairly decide who gets an organ transplant? *Harvard Business Review*. <https://hbr.org/2020/12/can-ai-fairly-decide-who-gets-an-organ-transplant>

Batty, M. (2020). Defining smart cities. In K. S. Willis & A. Aurigi (Eds.), *The Routledge companion to smart cities* (1st ed., pp. 51–60). Routledge. <https://doi.org/10.4324/9781315178387-5>

Bertuzzi, L. (2023, January 10). AI Act: MEPs want fundamental rights assessments, obligations for high-risk users. *EURACTIV*. <https://www.euractiv.com/section/artificial-intelligence/news/ai-act-mep-s-want-fundamental-rights-assessments-obligations-for-high-risk-users/>

Cobbe, J., Lee, M. S. A., Janssen, H., & Singh, J. (2020). Centering the law in the digital state. *Computer*, 53(10), 47–58. <https://doi.org/10.1109/MC.2020.3006623>

Council of Europe. (n.d.). *History of artificial intelligence*. Artificial intelligence. <https://www.coe.int/en/web/artificial-intelligence/history-of-ai>

Council of the European Union. (2022a). *Proposal for a Directive of the European Parliament and of the Council on corporate sustainability due diligence and amending Directive (EU) 2019/1937 – General Approach* (Institutional File 2022/0051(COD)).

Council of the European Union. (2022b). *Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – General approach* (Interinstitutional File 2021/0106(COD)).

de Andrade, N. N. G., & Kontschieder, V. (2021). *AI impact assessment: A policy prototyping experiment* [Report]. Open Loop. [https://openloop.org/wp-content/uploads/2021/01/AI\\_Impact\\_Assessment\\_A\\_Policy\\_Prototyping\\_Experiment.pdf](https://openloop.org/wp-content/uploads/2021/01/AI_Impact_Assessment_A_Policy_Prototyping_Experiment.pdf)

Donahoe, E., & Metzger, M. M. (2019). Artificial intelligence and human rights. *Journal of Democracy*, 30(2), 115–126. <https://doi.org/10.1353/jod.2019.0029>

Ebers, M., Hoch, V. R. S., Rosenkranz, F., Ruschemeier, H., & Steinrötter, B. (2021). The European Commission's Proposal for an Artificial Intelligence Act – A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *J, 4*(4), 589–603. <https://doi.org/10.3390/j4040043>

Eckhoff, D., & Wagner, I. (2018). Privacy in the smart city – Applications, technologies, challenges, and solutions. *IEEE Communications Surveys & Tutorials, 20*(1), 489–516. <https://doi.org/10.1109/CO MST.2017.2748998>

Edwards, L. (2016). Privacy, security and data protection in smart cities: A critical EU law perspective. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2711290>

Edwards, L. (2022). *Expert explainer: The EU AI Act: A summary of its significance and scope* [Report]. Ada Lovelace Institute. <https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf>

Edwards, L., & Veale, M. (2017). *Slave to the algorithm? Why a 'right to an explanation' is probably not the remedy you are looking for* [Preprint]. LawArXiv. <https://doi.org/10.31228/osf.io/97upg>

European A.I. Alliance. (2018). *Artificial intelligence impact assessment* [Report]. ECP Platform for the Information Society. <https://static1.squarespace.com/static/5b7877457c9327fa97fef427/t/5c368c611ae6cf01ea0fba53/1547078768062/Artificial+Intelligence+Impact+Assessment+-+English.pdf>

European Commission. (2020). *White paper on artificial intelligence. A European approach to excellence and trust* (White Paper COM(2020) 65 final; pp. 1–26). [https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust\\_en](https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en)

European Commission. (2021). *Proposal for a regulation of the European Parliament and of the council laying down harmonised rules on artificial intelligence (artificial intelligence act) and amending certain Union legislative acts* (COM(2021) 206 final 2021/0106(COD)). <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206>

European Commission. (2022). *Proposal for a Directive on corporate sustainability due diligence* (COM(2022) 71 final 2022/0051 (COD)).

European Commission. (n.d.). *Smart cities*. [https://ec.europa.eu/info/es-regionu-ir-miestu-pletra/temos/miestai-ir-miestu-pletra/miestu-iniciatyvos/smart-cities\\_en](https://ec.europa.eu/info/es-regionu-ir-miestu-pletra/temos/miestai-ir-miestu-pletra/miestu-iniciatyvos/smart-cities_en)

European Economic and Social Committee. (2016). *Artificial intelligence: The consequences of artificial intelligence on the (digital) single market, production, consumption, employment and society* (Own-Initiative Opinion INT/806-EESC-2016-05369-00-00-AC-TRA). European Economic and Social Committee. <https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/artificial-intelligence-consequences-artificial-intelligence-digital-single-market-production-consumption-employment-and>

European Parliament Committee on Legal Affairs. (2022). *Draft report on the proposal for a directive of the European Parliament and of the Council on Corporate Sustainability Due Diligence and amending Directive EU) 2019/1937 (COM(2022)0071 – C9-0050/2022 – 2022/0051(COD))*.

European Union Agency for Fundamental Rights. (2022). *Bias in algorithms: Artificial intelligence and discrimination* [Report]. European Union Agency for Fundamental Rights. <https://doi.org/10.2811/25847>

Floridi, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. *Philosophy & Technology, 34*(2), 215–222. <https://doi.org/10.1007/s13347-021-00460-9>

Forest Peoples Programme. (2022). *Forest Peoples Programmes' feedback on the European Commission*

*proposal for a directive on corporate sustainability due diligence*. Forest Peoples Programme. <https://www.forestpeoples.org/sites/default/files/documents/FFP%20feedback%20-%20CSDD.pdf>

Goodman, E. P. (2020). Smart city ethics: How “smart” challenges democratic governance. In M. D. Dubber, F. Pasquale, & S. Das (Eds.), *The Oxford Handbook of Ethics of AI* (pp. 822–839). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190067397.013.53>

Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45, 105681. <https://doi.org/10.1016/j.clsr.2022.105681>

Gryz, J., & Rojszczak, M. (2021). Black box algorithms and the rights of individuals: No easy solution to the “explainability” problem. *Internet Policy Review*, 10(2). <https://doi.org/10.14763/2021.2.1564>

Hallo de Wolf, A. (2012). *Reconciling privatization with human rights*. Intersentia.

Helbing, D., Fanitabasi, F., Giannotti, F., Hänggli, R., Hausladen, C. I., Van Den Hoven, J., Mahajan, S., Pedreschi, D., & Pournaras, E. (2021). Ethics of smart cities: Towards value-sensitive design and co-evolving city life. *Sustainability*, 13(20), 11162. <https://doi.org/10.3390/su132011162>

Henley, J. (2021, January 15). Dutch government resigns over child benefits scandal. *The Guardian*. <https://www.theguardian.com/world/2021/jan/15/dutch-government-resigns-over-child-benefits-scandal>

Herath, H. M. K. K. M. B., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, 2(1), 100076. <https://doi.org/10.1016/j.jjime.2022.100076>

Hesselman, M., Wolf, A. H., & Toebes, B. (Eds.). (2017). *Socio-economic human rights in essential public services provision*. Routledge. <https://doi.org/10.4324/9781315618081>

High-Level Expert Group Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. <https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html>

Hollands, R. G. (2008). Will the real smart city please stand up?: Intelligent, progressive or entrepreneurial?, *City*, 12(3), 303–320. <https://doi.org/10.1080/13604810802479126>

Hollands, R. G. (2015). Critical interventions into the corporate smart city. *Cambridge Journal of Regions, Economy and Society*, 8(1), 61–77. <https://doi.org/10.1093/cjres/rsu011>

Human Rights Council. (2021). *Report of the United Nations High Commissioner for Human Rights: The right to privacy in the digital age* (Report A/HRC/48/31).

International Federation for Human Rights. (2022). *Europe can do better: How EU policy makers can strengthen the corporate sustainability due diligence directive* (Position Paper No. 794a). International Federation for Human Rights. <https://www.fidh.org/IMG/pdf/duediligence.pdf>

Jiang, H., Geertman, S., & Witte, P. (2022). Smart urban governance: An alternative to technocratic “smartness”. *GeoJournal*, 87(3), 1639–1655. <https://doi.org/10.1007/s10708-020-10326-w>

Kempin Reuter, T. (2020). *Smart city visions and human rights: Do they go together?* (No. 2020–006; Carr Center Discussion Paper Series). Harvard Kennedy School. <https://carrcenter.hks.harvard.edu/publications/smart-city-visions-and-human-rights-do-they-go-together>

Kitchin, R. (2015). Making sense of smart cities: Addressing present shortcomings. *Cambridge Journal of Regions, Economy and Society*, 8(1), 131–136. <https://doi.org/10.1093/cjres/rsu027>

Kitchin, R., Cardullo, P., & Felicianantonio, C. D. (2018). *Citizenship, justice and the right to the smart city*

(Working Paper No. 41; The Programmable City). <https://doi.org/10.31235/osf.io/b8aq5>

Kitchin, R., Coletta, C., Evans, L., Heaphy, L., & MacDonncha, D. (2017). Smart cities, epistemic communities, advocacy coalitions and the 'last mile' problem. *It - Information Technology*, 59(6), 275–284. <https://doi.org/10.1515/itit-2017-0004>

Koops, B.-J. (2021). The concept of function creep. *Law, Innovation and Technology*, 13(1), 29–56. <http://doi.org/10.1080/17579961.2021.1898299>

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2016). Accountable algorithms. *University of Pennsylvania Law Review*, 165, 633–705.

Lacroix, C. (2020). *Preventing discrimination caused by the use of artificial intelligence* (Doc. 15151). Committee on Equality and Non-discrimination, Parliamentary Assembly, Council of Europe. [http://www.eerstekamer.nl/bijlage/20201105/preventing\\_discrimination\\_caused/document3/f=vldiexl467t0.pdf](http://www.eerstekamer.nl/bijlage/20201105/preventing_discrimination_caused/document3/f=vldiexl467t0.pdf)

Lane, L. (2022). Clarifying human rights standards through artificial intelligence initiatives. *International and Comparative Law Quarterly*, 71(4), 915–944. <https://doi.org/10.1017/S0020589322000380>

Lane, L. (2023). Artificial intelligence and human rights: Corporate responsibility in AI governance initiatives. *Nordic Journal of Human Rights*, 1–22. <https://doi.org/10.1080/18918131.2022.2137288>

Larson, E. J. (2021). *The myth of artificial intelligence: Why computers can't think the way we do*. Belknap Press.

Maas, M. M. (2019). International law does not compute: Artificial intelligence and the development, displacement or destruction of the global legal order. *Melbourne Journal of International Law*, 20(1), 29–57. <http://classic.austlii.edu.au/au/journals/MelbJIL/2019/3.html>

Mahler, T. (2022). Between risk management and proportionality: The risk-based approach in the EU's Artificial Intelligence Act Proposal. *The Swedish Law and Informatics Research Institute*, 247–270. <https://doi.org/10.53292/208f5901.38a67238>

Malaihollo, M. (2021). Due diligence in international environmental law and international human rights law: A comparative legal study of the nationally determined contributions under the Paris Agreement and positive obligations under the European Convention on Human Rights. *Netherlands International Law Review*, 68(1), 121–155. <https://doi.org/10.1007/s40802-021-00188-5>

McCorquodale, R., & Nolan, J. (2021). The effectiveness of human rights due diligence for preventing business human rights abuses. *Netherlands International Law Review*, 68(3), 455–478. <https://doi.org/10.1007/s40802-021-00201-x>

Monnheimer, M. (2021). *Due diligence obligations in international human rights law*. Cambridge University Press. <https://doi.org/10.1017/9781108894784>

Muller, C. (2020). *The impact of artificial intelligence on human rights, democracy and the rule of law* (Report CAHAI(2020)06-fin). Council of Europe, Ad hoc Committee on Artificial Intelligence.

NJCM et al. V The Dutch State (SyRI), C/09/550982 / HA ZA 18-388 (The Hague District Court 5 February 2020). <https://deeplink.rechtspraak.nl/uitspraak?id=ECLI:NL:RBDHA:2020:865>

Open-ended intergovernmental working group on transnational corporations and other business enterprises with respect to human rights. (2021). *Third revised draft legally binding instrument to regulate, in international human rights law, the activities of transnational corporations and other*

*business enterprises*. United Nations Human Rights Council. [www.ohchr.org/en/hrbodies/hrc/wgtran-scorp/pages/igwgontnc.aspx](http://www.ohchr.org/en/hrbodies/hrc/wgtran-scorp/pages/igwgontnc.aspx)

Organisation for Economic Co-operation and Development. (2011). *OECD Guidelines for Multinational Enterprises* (2011th ed.). Organisation for Economic Co-operation and Development. <http://mneguidelines.oecd.org/guidelines/>

Organisation for Economic Co-operation and Development. (2016). *OECD Due Diligence Guidance for Responsible Business Conduct*. Organisation for Economic Co-operation and Development. <http://mneguidelines.oecd.org/due-diligence-guidance-for-responsible-business-conduct.htm>

Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence* (Recommendation OECD/Legal/0449; Legal Instruments). Organisation for Economic Co-operation and Development. <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449>

Organisation for Economic Co-operation and Development. (2021). *AI in business and finance* (OECD Business and Finance Outlook). Organisation for Economic Co-operation and Development. [www.oecd.org/daf/oecd-business-and-finance-outlook-26172577.htm](http://www.oecd.org/daf/oecd-business-and-finance-outlook-26172577.htm)

Pellegrin, J., Colnot, L., & Delponte, L. (2021). *Research for REGL Committee: Artificial intelligence and urban development* [Study]. European Parliament, Policy Department for Structural and Cohesion Policies. [https://www.europarl.europa.eu/thinktank/en/document/IPOL\\_STU\(2021\)690882](https://www.europarl.europa.eu/thinktank/en/document/IPOL_STU(2021)690882)

Phillips, P. J., Hahn, C. A., Fontana, P. C., Broniatowski, D. A., & Przybocki, M. A. (2020). *Four principles of explainable artificial intelligence* [Preprint]. <https://doi.org/10.6028/NIST.IR.8312-draft>

Qarri, A., & Gill, L. (2022). *Smart cities and human rights* (Community Solutions Network) [Research Brief]. Future Cities Canada. [https://opennorth.ca/wp-content/uploads/legacy/RB\\_-\\_Human\\_Rights.pdf](https://opennorth.ca/wp-content/uploads/legacy/RB_-_Human_Rights.pdf)

Rachovitsa, A., & Johann, N. (2022). The human rights implications of the use of AI in the digital welfare state: Lessons learned from the Dutch *SyRI* Case. *Human Rights Law Review*, 22(2), ngac010. <https://doi.org/10.1093/hrlr/ngac010>

Ranchordás, S., & Goanta, C. (2020). The new city regulators: Platform and public values in smart and sharing Cities. *Computer Law & Security Review*, 36, 105375. <https://doi.org/10.1016/j.clsr.2019.105375>

Roig, A. (2017). Safeguards for the right not to be subject to a decision. *European Journal of Law and Technology*, 8(3). <https://ejlt.org/index.php/ejlt/article/view/570>

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206–215. <https://doi.org/10.1038/s42256-019-0048-x>

Ruggie, J. (2011). *Report of the special representative of the secretary-general on the issue of human rights and transnational corporations and other business enterprises* (Report A/HRC/17/31). United Nations Human Rights Council. <https://digitallibrary.un.org/record/705860>

Sawhney, N. (2022). Contestations in urban mobility: Rights, risks, and responsibilities for urban AI. *AI & SOCIETY*. <https://doi.org/10.1007/s00146-022-01502-2>

Scherer, M. U. (2015). Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2609777>



Schuett, J. (2022). *Risk management in the Artificial Intelligence Act* (arXiv:2212.03109). arXiv. <http://arxiv.org/abs/2212.03109>

Seldam, B., & Brenninkmeijer, A. (2021, April 30). The Dutch benefits scandal: A cautionary tale for algorithmic enforcement [Blog post]. *EU Law Enforcement*. <https://eulawenforcement.com/?p=7941>

Sengupta, U., & Sengupta, U. (2022). Why government supported smart city initiatives fail: Examining community risk and benefit agreements as a missing link to accountability for equity-seeking groups. *Frontiers in Sustainable Cities*, 4, 960400. <https://doi.org/10.3389/frsc.2022.960400>

Shanahan, M. (2015). *The technological singularity* (MIT Press Essential Knowledge Series). MIT Press. <https://mitpress.mit.edu/9780262527804/the-technological-singularity/>

Sharif, R. A., & Pokharel, S. (2022). Smart city dimensions and associated risks: Review of literature. *Sustainable Cities and Society*, 77, 103542. <https://doi.org/10.1016/j.scs.2021.103542>

Silver, D., Huang, A., Maddison, C. J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., Dieleman, S., Grewe, D., Nham, J., Kalchbrenner, N., Sutskever, I., Lillicrap, T., Leach, M., Kavukcuoglu, K., Graepel, T., & Hassabis, D. (2016). Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587), 484–489. <https://doi.org/10.1038/nature16961>

Simmons & Simmons. (2022, March 3). *ESG: Human rights and environmental due diligence proposal* [Publication]. Simmons+Simmons Insights. <https://www.simmons-simmons.com/en/publications/cl0b5oj4t27mo0b00yozvtgu7/esg-human-rights-and-environmental-due-diligence-proposal>

Sovrano, F., Sapienza, S., Palmirani, M., & Vitali, F. (2022). Metrics, explainability and the European AI Act Proposal. *J*, 5(1), 126–138. <https://doi.org/10.3390/j5010010>

Umbrello, S., & Van De Poel, I. (2021). Mapping value sensitive design onto AI for social good principles. *AI and Ethics*, 1(3), 283–296. <https://doi.org/10.1007/s43681-021-00038-3>

United Nations Committee on Economic, Social, and Cultural Rights. (1999). *General Comment No. 12: The Right to Adequate Food (Art. 11 of the Covenant)* (E/C.12/1999/5).

United Nations Committee on Economic, Social, and Cultural Rights. (2017). *General comment No. 24 on State obligations under the International Covenant on Economic, Social and Cultural Rights in the context of business activities* (E/C.12/GC/24).

United Nations Committee on the Rights of the Child. (2021). *General Comment No 25 on children's rights in relation to the digital environment* (CRC/C/GC/25).

United Nations Educational, Scientific and Cultural Organization. (2021). *Recommendation on the ethics of artificial intelligence* (SHS/BIO/PI/2021/1).

Veale, M., & Zuiderveen Borgesius, F. (2021). Demystifying the Draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4), 97–112. <https://doi.org/10.9785/cr-2021-220402>

Vilone, G., & Longo, L. (2021). Notions of explainability and evaluation approaches for explainable artificial intelligence. *Information Fusion*, 76, 89–106. <https://doi.org/10.1016/j.inffus.2021.05.009>

Voorwinden, A. (2021). The privatised city: Technology and public-private partnerships in the smart city. *Law, Innovation and Technology*, 13(2), 439–463. <https://doi.org/10.1080/17579961.2021.1977213>

Wachter, S. (2022). The theory of artificial immutability: Protecting algorithmic groups under anti-discrimination law. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.4099100>

Whittlestone, J., & Avin, S. (n.d.). *Why the AI impacts ecosystem must move beyond 'near-term' and 'long-term'*. University of Cambridge Trust & Technology Initiative. <https://www.trusttech.cam.ac.uk/perspectives/sector-specific-applications/why-ai-impacts-ecosystem-must-move-beyond-near-term-and>

Yampolskiy, R. V. (2019). *Unpredictability of AI* (arXiv:1905.13053). arXiv. <http://arxiv.org/abs/1905.13053>

Published by



ALEXANDER VON HUMBOLDT  
INSTITUTE FOR INTERNET  
AND SOCIETY

in cooperation with



CREATE



centre  
— internet  
et — society



R&I IN3  
Internet  
interdisciplinary  
Institute  
Universitat Oberta de Catalunya



UNIVERSITY OF TARTU  
Johan Skytte Institute of  
Political Studies