Delft University of Technology

AI for crisis decisions

Comes, Tina

DOI
10.1007/s10676-024-09750-0

Publication date
2024

Document Version
Final published version

Published in
Ethics and Information Technology

Citation (APA)
Comes, T. (2024). AI for crisis decisions. *Ethics and Information Technology*, *26*(1), Article 12.
https://doi.org/10.1007/s10676-024-09750-0

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

Ethics and Information Technology (2024) 26:12
https://doi.org/10.1007/s10676-024-09750-0

ORIGINAL PAPER

AI for crisis decisions

Tina Comes[1]

© The Author(s) 2024

Abstract

Increasingly, our cities are confronted with crises. Fuelled by climate change and a loss of biodiversity, increasing inequalities and fragmentation, challenges range from social unrest and outbursts of violence to heatwaves, torrential rainfall, or epidemics. As crises require rapid interventions that overwhelm human decision-making capacity, AI has been portrayed as a potential avenue to support or even automate decision-making. In this paper, I analyse the specific challenges of AI in urban crisis management as an example and test case for many *super wicked* decision problems. These super wicked problems are characterised by a coincidence of great complexity and urgency. I will argue that from this combination, specific challenges arise that are only partially covered in the current guidelines and standards around trustworthy or human-centered AI. By following a decision-centric perspective, I argue that to solve urgent crisis problems, the context, capacities, and networks need to be addressed. AI for crisis response needs to follow dedicated design principles that ensure (i) human control in complex social networks, where many humans interact with AI; (ii) principled design that considers core principles of crisis response such as solidarity and humanity; (iii) designing for the most vulnerable. As such this paper is meant to inspire researchers, AI developers and practitioners in the space of AI for (urban) crisis response – and other urgent and complex problems that urban planners are confronted with.

Keywords Human-AI interaction · Decision theory · Crisis management · Human-centred AI · Responsible AI

Introduction

Our cities are complex social-technical systems in which people continuously interact with infrastructures and technologies as they travel, work, or go shopping. Confronted with climate changes, cities around the world are under pressure to reach sustainability targets, and transform their infrastructural systems, ranging from the energy transition to urban gardening for local food production or electric vehicles and ride sharing. Complex systems like our cities are always difficult to control and steer, especially during transitions.

A major challenges is that our transitioning cities are increasingly confronted with crises: the 6th Assessment report of the IPCC once more issues stark warnings: climate change is confronting our cities with an increasing number and severity of extreme events, ranging from extreme heatwaves to heavy precipitation, rising see levels and droughtsPersson and Tinghög (2023); (Seneviratne et al., 2021) sho In their 2023 Global Risks report, the World Economic Forum refers to the global *"new normal*" as a situation where in our cities basic needs, such as food, energy, and security, remain often unmet (WEF, 2023). Combined with the increasing interconnectedness of socio-technical systems this trend implies that decision-makers need to make increasingly complex and far-reaching decisions at an accelerating pace. In other words: crises fundamentally transform the nature and timescapes of planning and decision-making: the proliferation and volatility of information (Höchtl et al., 2016; Tsoukias et al., 2013) combined with the urgent need to alleviate human suffering pressure decision-makers to rapidly respond to the increasing number of crises and disasters worldwide.

Recognizing this great complexity of (urban) crisis response, AI is frequently portrayed as a way ahead to improve rapid information collection, analysis, and decision support. Facilitated by advances in (remote) sensing and crowd sourcing, increasingly available open data

✉ Tina Comes
t.comes@tudelft.nl

[1] Faculty Technology Policy & Management, TU Delft, Delft, The Netherlands

Published online: 14 February 2024

and accelerating computational capacities, the hope is that AI can help humans to better prepare for, respond to and recover from crises (Sun et al., 2020). At the same time, reliance on AI - and the introduction of digital technology more broadly - have added a layer of interdependencies, and therefore introduced additional vulnerabilities. How, for instance, can we live in smart homes during a power or internet outages? Sure enough, critiques of AI have pointed to the potential downsides of experimentation with AI and digital technology in a very vulnerable context (Sandvik et al., 2017); the need for accountability and the lack of adequate legal frameworks (Coppi et al., 2021); as well as the tremendous carbon footprint of AI, especially for training machine learning models (Van Wynsberghe, 2021).

To guide the use of AI amidst the promises and perils, a series of guidelines, standards and principles have been put forward. Some of the most prominent ones include guidelines by regulatory bodies such as the United Nations Educational Scientific and Cultural Organization (UNESCO) recommendations on the Ethics of AI (UNESCO, 2022), the Organisation for Economic Co-operation and Development (OECD) Recommendation of the Council on Artificial Intelligence (OECD, 2019), the European Commission's recommendations by the High-Level Expert Group on AI (EC, 2019) or the Institute of Electrical and Electronics Engineers' (IEEE) standards for Ethically Aligned Design of Autonomous and Intelligent Systems (IEEE, 2019). For an overview, see (Jobin et al., 2019). All of them embrace the idea of an AI that needs to be designed to be explainable, accountable, trustworthy, and fair – corresponding to the central tenets of human-centred AI (Shneiderman, 2020). Human-centered AI should be "*amplifying, augmenting, and enhancing human performance in ways that make systems reliable, safe, and trustworthy*" (Shneiderman, 2020). However, putting forward fundamental principles leaves many questions about the concrete design requirements and implementation (Mittelstadt, 2019).

Therefore, if the aim is to design a human-centered AI that supports, facilitates or even makes decisions in crises, an understanding human decision-making is crucial (Miller, 2019). Decision theory, as a field, is concerned with both human decision behaviour, cognition and interaction (descriptive) as well as with how optimisation and how we should make decisions (normative) (Tsoukiàs, 2008). Importantly, decision theory stresses the prominence of the context (French & Geldermann, 2005), which also finds an echo in the calls for 'contextual transparency' for automated systems (Sloane et al., 2023). Especially the context of crises is crucial, since crises have been shown to change human information processing and decision behaviour (Klein et al., 2010; Paulus et al., 2022; Weick et al., 2005). Nevertheless, decision theory has remained largely disconnected from the discourse around human-centered or explainable AI (Främling, 2020). This oversight leaves a gap in our understanding of the interactions between humans and AI (Rahwan et al., 2019), and unclear or inadequate design requirements (Mittelstadt, 2019).

To start exploring this gap, this paper aims to establish a research agenda around the decision-theoretical implications for AI in the context of crisis response. The underlying research question is: *what are research directions for a decision theoretical approach to support the design and development of human-centred AI for crisis management?* As indicated before, contextualisation is central to both decision theory and AI. Therefore, this article first explores the context of crisis decisions as super-wicked problems before unpacking the specifics of AI for crisis response. By taken a decision-theoretical perspective, I will analyse the implications of increasingly automating information acquisition, analysis, or decision-making. From there, I will highlight the differences between crisis and conventional planning decisions and argue for dedicated design principles for AI that are tailored for crises. The paper concludes with a discussion of important research directions for crisis management as an example case of super-wicked urban decisions.

Crises as super-wicked decision problems

Why is it worthwhile to analyse decision-making in crises, in addition to the many problems that decision-makers are already confronted with? In this section, I will argue that crisis decisions are an ideal testing ground to understand the intricate feedback between human decision-making and AI-systems in urban contexts, given the time pressure and the value-laden context. While the Covid-19 pandemic has fueled interest in urban resilience (Champlin et al., 2023), there remains limited research on crisis management in the smart city – for a recent review, see (Alshamaila et al., 2023). Generally, the focus on the smart cities crisis literature is exploiting sensing technologies for real-time decision support (Wang & Li, 2021; Yang et al., 2017), but there is limited reflection on how people interact with technology in the smart city.

In his article about AI and smart cities, Batty (2018) argued that many of the challenges around the smart city are "*peculiarly human*", or intrinsically related to human reasoning and the dilemmas or *"hard choices"* that planners need to make. The combination of urgency and complexity that we observe in crises is characteristic for many super-wicked problems that urban planners are confronted with (Levin et al., 2012), ranging from climate change to sustainability or migration. Such urgent and complex problems are known to change information processing and decision behaviour (Weick et al., 2005). So, if AI in smart cities is indeed promising to address problems in healthcare, mobility, or energy (Herath & Mittal, 2022), then I argue that analysing (AI for) urban crisis management can help us make headway in the crucial area of understanding the interaction of human decision-makers and machines.

Decisions in (urban) crises are complex since crises affect virtually all aspects of our societies (Comes et al., 2022). A crisis occurs when people perceive a severe threat to the fundamental values or functioning of a society or system, requiring an immediate response despite the prevailing uncertainties (Boin et al., 2016; Rosenthal et al., 1989). Even though many taxonomies exist that distinguish crises according to their scale, the underlying hazard (natural vs. human-made), time (sudden vs. slow onset) or geographical scale (from local to global), traditionally, crisis management has focused on actual or proverbial firefighting in a clearly defined region or sector. Not surprisingly then, also design of coordination mechanisms and decision support systems or AI to support crisis management has focused primarily on providing rapid information to people in various – yet shifting – roles (Comfort, 2007; Quarantelli, 1988; Turoff et al., 2004).

Today, driven by the proliferation of AI, crises are harder to manage as they evolve within complex and deeply interconnected systems (Renn & Lucas, 2021). This implies that crises increasingly cascade into distant regions or other sectors (Boin, 2019). Amplifying feedback loops at different timescales may eventually result in largescale disruptions (Helbing, 2009). For instance, the war in Ukraine has caused food insecurity and inflation across the globe (Arndt et al., 2023). The response to the Covid19 pandemic, initially thought of as a 'health crisis', has led to a loss of income, food insecurity, limited access to education or the ability to access and purchase basic necessities – especially in low-income countries (Josephson et al., 2021).

Further, AI and digital technology act as an amplifier of the emergent behaviour of individuals that can fundamentally impact the impact of crisis decisions. From fear to 'pandemic fatigue' driving non-compliance to Covid19 restrictions and political discontent (Jørgensen et al., 2022) to the outpour of solidarity and help after natural disasters such as the floods that affected Western Europe in 2021 (Bier et al., 2023), the spontaneous responses by self-organizing individuals, emergent or organized voluntary groups are a common feature of crises, especially in urbanized areas (Twigg & Mosel, 2017). And there are many calls to localize response and leverage self-organization to create participatory resilience (Mahajan et al., 2022; Nespeca et al., 2020).

The initial phase of a crisis response is often heavily resource-constrained, leading to a combination of uncertainty, cognitive and moral overload and fragmented sensemaking (Comes et al., 2020; Ishmaev et al., 2021). Here, sensemaking is the cognitive process by which decision-makers interpret the situation, and constructing meaning from data or experiences in order to understand and respond to complex, ambiguous, or novel situations (Weick et al., 2005). The messy characteristics of a crisis pose a double challenge (Paulus et al., 2022) (a) data may be unavailable, uncertain, conflicting or biased, given limited access or data collection regimes – with limited options to collect additional data because of the time constraints; and (b) the cognitive processes of analysis or decision-makers is under strain, given the urgency and high stakes of the situation, leading to biased decisions.

Despite the highly volatile, complex, and uncertain environment, the time pressure of crises forces decision-makers to make choices that have far reaching consequences in a much shorter time than traditional models require (French & Geldermann, 2005). Time is of the essence in urban search and rescue (Kleiner & Dornhege, 2007): after the earthquake that hit Turkey and Syria in spring 2023, the world followed the desperate race of urban search and rescue teams against time, and their attempts to locate and reach victims under the collapsed buildings and infrastructures and bring them to safety. Logistics and planning in these humanitarian crises are time-compressed (Holguín-Veras et al., 2012) On the other end of the spectrum, Covid19 has brought about a rapid fire of restrictions and measures ranging from tracking apps to lockdowns, from travel bans to large scale investment programs – all of which were made rapidly despite having a profound impact that was at times poorly understood (Atkinson et al., 2020; Sigala et al., 2022). The war in Ukraine led to fast measures to compensate the lack of Russian gas by a turn to coal (Aitken & Ersoy, 2023), despite its potential impact on greenhouse gas targets. And while conventionally urban planning decisions are made in structured planning processes that take years, in the recovery phase to an urban disaster, planning is focused on rapidly rebuilding the status ex ante, even though it is this very status that may have led to the disaster in the first place (Krishnan et al., 2024).

These decisions have in common that despite their profound impact on human lives and livelihoods, the conventional time frames to make decisions with far reaching consequences collapses. We observe fundamentally different timescapes of decision-making as highlighted in Fig. 1. For conventional problems in planning and decision-making (in blue), the further the decisions reach – and the more uncertain therefore is related to future predictions – the longer the time to make the decision corresponds. For crisis decisions (red in Fig. 1), however, the window to decide dramatically shrinks. Since inaction or waiting until further information

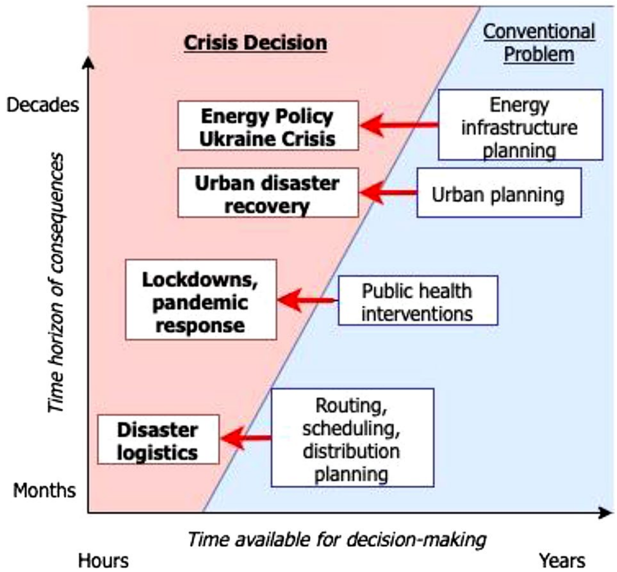


Fig. 1 The shifting timescapes of decision-making

becomes available is not possible, decision-makers need to rapidly act. This leads to a time compression of complex problems with far-reaching consequences (red arrows).

In sum, decision-makers in crises are confronted with a combination of complexity and urgency that is typical for super-wicked problems (Levin et al., 2012). The phenomenon of complexity and time compression is further exacerbated characterised by decision density, by which – especially in the early phases of a crisis – an exceptional number of decisions need to be made rapidly (Baharmand et al., 2019).

AI in crisis decisions

Artificial Intelligence is becoming increasingly prominent as a vehicle to support or replace human decision-making in crises and disasters. With the increasing prominence of AI, there is also a plethora of definitions that aim to characterise the behaviour or function of what constitutes an 'AI' (Krafft et al., 2020). Since this research is situated at the interface of science and policy, I follow here the OECD definition "*An AI system is a machine-based system that can* […] *make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy*." (OECD, 2019).

There is a broad range of applications from automating early warning to damage detection and response and recovery planning. Often, AI systems are used to improve the predictive accuracy or speed of prediction as compared to traditional models (Guikema, 2020) to predict hazards, model human mobility or analyse supply chains. There are recent reviews that provide a full account of the different applications for crises and disasters (Munawar et al., 2022; Sun et al., 2020) or risk communication (Ogie et al., 2018), therefore I provide here only a short overview of selected applications, examples and data requirements in Table 1. Applications are mapped on the crisis management cycle of preparedness, response, and recovery to indicate the different timescales and actors involved in the application of the AI system.

Human control and oversight are at the heart of human-centred AI (Shneiderman, 2020) and instrumental to the discussions around AI, see e.g., the UNESCO guidelines (UNESCO, 2022) or the EU's guidelines on trustworthy AI (EC, 2019). Yet, increasingly there is a push for automation in crisis management (Coppi et al., 2021). Therefore, Table 1 includes automation. Automation can be defined as "*a device or system that accomplishes (partially or fully) a function that was previously, or conceivably could be, carried out (partially or fully) by a human operator*" (Parasuraman et al., 2000). While colloquially, automation is understood as 'the machine taking over' a task at which it is 'better' (Bradshaw et al., 2012), there is more nuance to the degree of automation and control. One of the first frameworks to discuss automation was established by Sheridan et al. (1978) to understand the control of underwater vessels. Later, their work was refined to study automation of air traffic control, resulting in a 10-point scale to measure automation of decision making that describes the interaction between humans and machines in decision-making. The scale reaches from no interference of a computer via the computer making suggestions, to the computer ignoring - not even informing - the human (Council, 1998). Building on these initial ideas, Parasuraman et al. (2000) put forward a framework that establishes four axes of autonomy or control related to the full cycle of decision-making: (1) information acquisition, (2) analysis of information, (3) decision or choice of action, and (4) execution of action. From there, Endsley (2017) makes the link to situational awareness stressing that automation can actually impede situational awareness by adding a layer of complexity and the need to continuously monitor the machine. Thereby, instead of increasing efficiency and effectiveness, automation actually leads to lower performance (Strauch, 2017). To mitigate the effects, *adjustable autonomy* refers to a system where the level of automation can be dynamically adapted to the context (for a review, see (Mostafa et al., 2019).

However, all these automation studies assume – often implicitly - that there is a single decision-maker or operator, guiding a vessel or an aircraft. However, the pervasiveness of AI has introduced more decentralized or hybrid cases. For instance, if a script is used to acquire and download data from Twitter, it is obviously the machine that acquires data,

Table 1 Examples of applications of AI in crisis and disaster management

Application	Disaster Phase	Examples	Data source (examples)	Automation
Early Warning	Preparedness	Earthquakes (Grasso et al., 2007), flash floods (Cools et al., 2012)	Sensors, weather data	Information acquisition [hybrid]; information analysis; decision; execution
Forecast-based financing	Preparedness	Forecast-based early action (Wilkinson et al., 2018);	Climate, vulnerability & risk, weather, conflict, and socio-economic data	Information acquisition [hybrid]; information analysis; decision; execution
Famine declaration	Preparedness	Food security prediction (Martini et al., 2022)	Climate, vulnerability, weather, conflict, and population	Information acquisition [hybrid]; information analysis
Demand forecasting; accessibility	Response (sudden & slow onset)	Haiti earthquake (Gao et al., 2011); Queensland flooding (Kankanamge et al., 2020)	Social media	Information acquisition [hybrid]; information analysis
Urban crowding	Response (sudden onset)	Crowd control (Kumar & Parikh, 2023); police robots (Szocik & Abylkasymova, 2022)	Mobility data, social media	Information acquisition [robots]; information analysis; decision; execution [robots]
Damage and loss detection; environmental degradation	Response (sudden & slow onset), Accountability	Crisis information (Voigt et al., 2007); crop losses in Ukraine (Deininger et al., 2023)	Satellite imagery	Information acquisition; information analysis
Location and identification	Response (sudden onset)	Victim identification (Atif et al., 2021); urban search and rescue (Rizk et al., 2021)	Aerial imagery (e.g., UAVs)	Information acquisition; information analysis; decision [recommendation]
Migration management	Response (protracted)	Asylum determination and refugee tracking in Turkey (Elebe & Kurnaz, 2023)	Biometric information, personal data, tracking	information analysis; decision [recommendation]
Supply Chain	Response and recovery	Targeting vulnerable populations (Cash & voucher assistance) (Tschunkert & Vogel, 2023); Predicting vaccine supply chains (Ussher et al., 2021)	Consumption or vaccination patterns; personal information	Information analysis; decision [recommendation]

but the data itself is provided and fed into the machine by thousands of humans. Similarly, the execution of an action may require the interplay of humans and machines, for instance in the case that unmanned aerial vehicles (UAVs) are used to transport vaccines or medication to those in need. The '*many hands problem*' provide a framework to think about whom to attribute responsibility to – also in the context of AI (Coeckelbergh, 2020). The challenge here is the double complexity and emergence: in the context of a crisis, the roles and responsibility of human actors emerge, and are continuously adjusted to the situation at hand (Mendonça et al., 2007; Turoff et al., 2004). Technology and AI need to be designed to be adaptive to both the changes in roles and tasks, as well as to the situational and cognitive context of (human) decision-making to decide what needs to be assigned to whom in which role and in what situation exactly. Predefining all the possible contexts and requirements then become a challenge. The potential irreversibility of action, the high stakes and moral values at play further complicate the situation.

To address this complexity, human-centered AI suggests a variety of design principles to ensure to combine high levels of automation with human control (Riedl, 2019; Shneiderman, 2020). However, these design principles are not tailored to the crisis management context. For instance, the need for "incremental, and reversible action" (Shneiderman, 2020) may not be achievable in the context of crises where many actions lead to irreversible consequences (Pauwels et al., 2000).

Even though Table 1 just provides selected examples of uses of AI for crisis management, it shows applications across all phases of the crisis management cycle, ranging from preparedness to response and recovery. Data sources combine from contextual data (e.g., socio-demographics, climate, or risk profiles) and data about the impact of the hazard (e.g., aerial imagery) or the reaction of the population to it (e.g., mobility or social media data). By its very nature, contextual data is rather static, whereas data about the evolving disaster is highly volatile. Importantly, a recent review has shown that – driven by the differences in available data – there is a lack of standardization in terms of data sets, even for the very same problems (Casali et al., 2022), leading to a lack of comparability and standards. At the same time, for wicked and complex problems it is – by definition – impossible to guarantee that a model is comprehensive and captures all required variables. For instance, (Martini et al., 2022) excluded pandemics from their food security assessment – despite just having lived through the raging impacts of the Covid-19 pandemic. The need to a priori collect data and train the models may hamper their applicability and contextualization in crises.

For automation, not surprisingly, the focus is on automating information analyses, which is common to all examples. Further, AI is used to automate the acquisition of large amounts of volatile – or Big – data. This includes aerial and satellite imagery collected by drones; the using police robots to sense and control traffic; or sensors to detect and predict earthquakes. Hybrid examples include social media data as well as the combination and collation of different human and machine acquired data sets as is the case for early warnings.

Further, Table 1 presents two distinct cases for the automation of decision-making and implementation. First, as may be expected, the time pressure is the driver behind the use of AI and the push for fully automated decisions, spanning the full cycle of information acquisition to implementation. As there are conventionally only seconds between the detection of an earthquake and its onset. Similarly, the break of a levee or a nearing flash flood require extremely rapid intervention. Second, in other cases, such as forecast-based financing and early action, or automated decisions of food insecurity (and declaration of famine), the main argument to push for an AI is to make the decision more transparent and replicable, thereby removing it from the realm of political decisions and motivational biases (Lentz et al., 2019). Here, the 'black box' nature of political decisions is replaced by potential 'black box' AI models. These models also present a way to escape the cognitive (Van de Walle et al., 2016) or moral overload (Ishmaev et al., 2021) that is typical for crises.

To be sure, the move to an automated decision support system shifts the power towards the design of the protocols, and the many decisions about the data sets, algorithms, or thresholds that trigger assistance. However, in crises, we are not just 'replacing' the bias of a human by a biased machine. Rather, we need to take into account the interaction of humans with the machines and consider important path-dependencies that might emerge from this interaction. Research on the interaction of humans with information has shown that the very push for more 'evidence-based' decisions amplifies political biases (Paulus et al., 2023), and that there are strong path-dependencies by which initial biased decisions are not corrected (Paulus et al., 2022).

AI for crisis decisions? Towards decision-centred AI

Designing for crisis decisions – contextualised AI

Understanding the differences between crisis decision-making and conventional complex planning (e.g., in climate adaptation) needs to build on the recognition that planning and decision-making is increasingly a socio-technical process that AI systems. A social network of stakeholders and decision-makers from various organisations (e.g., NGOs, volunteers, police, fire fighters, businesses) make choices (e.g., if and where to evacuate, how to behave in a city in extreme heat, or how to accommodate a surge of incoming refugees) that depend on the availability and accessibility of urban infrastructures (e.g., hospitals, transportation), and increasingly depend on artificial intelligence (e.g., traffic jam predictions, or forecasted extreme events).

As diverse as the impacts of crisis decisions are therefore the actors and stakeholders involved. From professional crisis managers (e.g., police, fire fighters, health professionals) to policy-makers; from citizens and to specific professional interest groups or NGOs – the broad impact of choices that are made in crises implies any actors with different interests and preferences that have to be aligned, ranging from the (potentially) affected population to private and public sector, to the many volunteers that often engage in crisis response. While it may be difficult to activate stakeholders in the preparation for a crises, response will flip the situation into a messy multi-actor context characterised by ambiguity, fragmentation and emergence (Nespeca et al., 2020; Wolbers et al., 2018). The crux in crises is the interplay of professional teams with the affected actors and stakeholders that self-organize and act to respond to a crisis (Mahajan et al., 2022), where the actions, roles and responsibilities of individuals can therefore not be predetermined (Turoff et al., 2004). Because of the lack of pre-established rules, norms or processes that bind all actors and groups, information (sharing) becomes central to make sense of and coordinate in crises (Bharosa et al., 2009; Nespeca et al., 2020). In essence, information allows different actors to form a shared mental model of the situation and to translate this mental model via sensemaking trajectories into activities and tasks. While much focus in the discussion around the application of AI in crises evolves around automation of decisions, we also need to recognize that if an AI that steers or influences information sharing or analysis, it will have an influence on human sensemaking trajectories and network.

Therefore, we need to reconsider the current focus of much of the work on human autonomy to integrate the impact of automation on the broader sensemaking or meaning-making process. Rather than focusing on the ability of machines to carry out a specific task such as identifying damage or triggering financial assistance, we also need to understand in how far the automation of information collection and analysis leads to a removal of the emergency management authorities from context. van Wynsberghe and Comes (2020) point out by using the example of drones that such an automation step "*may push to de-contextualize care, also threaten to de-skill aid workers*".

To analyse the challenges for decision-making and the use of AI, we synthesize here various frameworks on decision-making in crises and disasters (Comes et al., 2011; French & Geldermann, 2005; Holguín-Veras et al., 2012; Paulus et al., 2022). We consolidate these characteristics with the value-driven character of crises that has been emphasized throughout the paper. While much literature on disaster and ethics primarily focus on health crises and duty of care (Leider et al., 2017), here we stress the need for principles around information and AI that have been forward by the UN (Van de Walle & Comes, 2015). We combine these aspects into a framework that summarizes the key differences between crisis decisions and complex planning (Table 2). The framework distinguishes nine dimensions: (i) uncertainty, (ii) stakes, (iii) complexity, (iv) cognitive load, (v) environment, (vi) actors and social networks, (vii) objectives, (viii) principles that guide the decision logic, and (ix) timescapes.

In a nutshell, crisis decisions combine the great complexity and uncertainty that is typical for planning with emergent and shifting objectives, volatile social networks, and great time pressure. This time pressure induces or amplifies cognitive and moral overload: in situations that combine the complexity with urgency, human decision-makers are far from rational and discount important cues, especially the distributive and long-term implications of their choices. Previous research has indeed shown that in such situations, decision-makers tend to focus on operational and local decisions that have an immediate impact on their environment, and tend to discard the long-term, strategic implications of their decisions (Comes et al., 2020).

If we design AI that increasingly automates information collection, processing, and decision-making, there are two major challenges: first, as discussed, human decision-making behaviour changes, and thereby also the interaction with the AI. Therefore, automated AI systems must be tuned to the changes in information processing and enactment that occurs in crises. Second, Table 2 shows that crisis management authorities are called to abide to specific principles and values, ranging from solidarity to humanity. However, given the lack of a clear operationalization of many of these principles, it is unclear how these principles operate if decision-making is automated (Coppi et al., 2021).

From principles to trade-offs

Crisis decisions affect the lives and livelihoods of many. Because of this broad impact, decisions in crises often cause moral dilemmas (Crawford & Finn, 2015; Qadir et al.,

Table 2 Characteristics of crisis decision-making

	Crisis decision-making	Conventional planning and decision-making
Uncertainty	Data and projections / forecasts (short- and long-term)	Projections / forecasts (long-term)
Stakes	Very high	Very high
Complexity	High	High
Cognitive load	Information and moral overload despite lacking and conflicting data	Informational overload
Environment	volatile	assumed to be stable (following trends / scenarios)
Actors and social networks	Emergent, multi-actor	Predefined or slowly shifting, multi-actor
Objectives	Emergent and volatile	Predefined, e.g., economic, environmental, distributional impacts
Guiding principles	Solidarity, humanity, neutrality, impartiality, independence	Trustworthiness, fairness, transparency, accountability, inclusivity
Timescapes	Urgent, delays not possible, immediate information decision feedback and accountability	Long-term, delays possible, delayed information decision feedback and accountability
Synthesis	**Urgency** + Complexity + Values	Complexity + Values

Note. The bold signifies the key differences between crisis and the 'usual' planning steps

2016). In the broadest sense, the well-known '*do no harm'* principle that guides crisis response can be viewed as an analogue to the principle of human beneficence and the need to mitigate risks that is widely promoted in the AI community (IEEE, 2019). Acknowledging that crisis decisions are always implicitly or explicitly guided by values, the European Working Group on Ethics in Science and New Technologies has advocated to prioritise the value of *solidarity* as a lens and guideline for ethical, and therefor also politically and socially more acceptable decisions (Prainsack et al., 2023). In contrast, in the humanitarian domain, the principles of *humanity*, *neutrality*, *impartiality*, and *independence* are thought to be the fundamental values to guide all information processing and decision-making (UN-OCHA, 2016). However, these principles are not conventionally considered in the literature on trustworthy or human-centred AI, which rather focuses on features such as accountability, explainability and transparency (Riedl, 2019). While these are undoubtedly crucial for AI in crisis response, what is missing is a discussion on what it will mean for an AI to uphold a value like solidarity or a principle like humanity. How can these principles be translated into protocols for data acquisition, or into an algorithm for information analysis or decision-making?

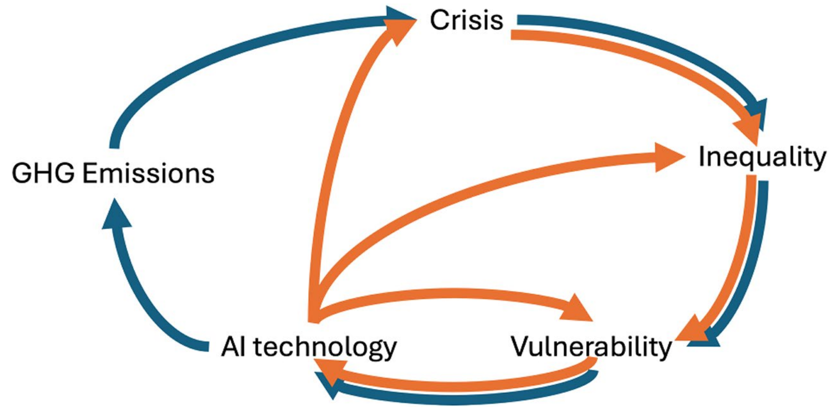
The principles that guide crisis response do not fall under the routinely used economic logic of costs and benefits that is typically used for planning. Even though broad welfare economics principles or distributive justice (via proxies such as access or distribution of wealth) are now increasingly advocated for in the realm of crisis response and resilience (de Bruijn et al., 2022; Holguin-Veras et al., 2013), these obviously do not correspond to principles such as 'solidarity' or 'humanity'. Because we are missing frameworks to translate these principles into machine readable formats, recommendations by an AI may be misleading and therefore discarded or lead to potentially harmful choices.

Further, crises also may bring about trade-offs between high-level principles that have been developed. Mittelstadt (2019) already points to the challenge that "*principles do not automatically translate to practice*", arguing for the need to develop methodologies to design concrete requirements and practices. However, the underlying assumption still seems to be that all principles can *all* be always maintained. Undoubtedly, the hallmark if crisis response is the *do no harm* principle (van Wynsberghe & Comes, 2020), which is translated into the *beneficence* of the technology in AI recommendations (UNESCO, 2022). However, it is not clear what the relation is of this principle with others, or how potential trade-offs should be made. For instance, in Covid19, the argument for increased surveillance (and potentially exposing vulnerable populations) was the safety and health of the population. Other questions include: would we sacrifice transparency or explainability for rapid action? And in how far would we relax accountability (as for the Covid19 vaccines) if we could potentially alleviate human suffering via an AI? Importantly, humans have been shown to change their preference for specific values under time pressure. For instance, Persson and Tinghög (2023) show that time pressure induces a preference for egalitarian over efficient solution. However, an AI may not be adaptive to the context. Therefore, what is needed are methodologies to decide a priori which principles dominate, and what may need to be sacrificed, and under which circumstances. Here, decision theory provides empirical methods to understand how humans make trade-offs (Keeney, 1996), and also analytical methods to formalize preferences and moral trade-offs into machine-readable formats by using preferential models (Holguin-Veras et al., 2013) or methods such as taboo trade-offs (Hanselmann & Tanner, 2008).

All in all, the urgency of a crisis clearly brings about challenges for the ethical design and oversight of an AI: crisis are often a space for experimentation (Sandvik et al., 2017). Given the urgency to alleviate human suffering, however, risk standards can shift, and what is acceptable in a crisis, and to whom, changes. An open question is: how can the tools that are rapidly developed and deployed in response to a crisis be tested and validated (Tzachor et al., 2020)? And who should set the guidelines and thresholds of what is acceptable, and to whom?

These questions also relate to the context of smart cities. According to Batty et al. (2012)'s seminal paper, smart cities entail both the planning, control and optimisation as well as "*technologies that ensure equity, fairness and realise a better quality of city life [and] ensure informed participation and create shared knowledge for democratic city governance*" (Batty et al., 2012). So far, however, in the development of AI for crisis management in the smart city, the focus has been overwhelmingly on sensing, analysing, and optimizing the response to crises (Yang et al., 2017). Limited attention has been paid to the potential downsides of the turn towards AI, such as introducing new vulnerabilities (Kitchin & Dodge, 2020), or the potential amplification of injustices (Kong & Woods, 2018). Because crises are known to especially affect the most vulnerable populations, it is critical to integrate the aspects of equity, fairness, participation into the design of technology for crisis management in the smart city. With respect to the sustainability of the technology (Van Wynsberghe, 2021), addressing the different timescapes of decision-making (cf. Table 2) is crucial: as crisis decisions tend to focus on the here and now, they tend to neglecting the contribution of the technology used to solve one crises to amplifying the climate crisis. Figure 2 highlights these interdependencies: on the short term both AI and crises are a driver of inequalities,

Fig. 2 Interdependencies of AI in the smart city



increasing vulnerability a thereby leading to new crises and calls for (AI) technology. At the same time, AI may introduce more technology-dependence, complexity and vulnerability (orange loops). On the longer term (blue loop), the combination of increased vulnerability combined with the emissions of unsustainable AI, both effects may drive future crises.

I argue that to move forward, a decision-theoretical perspective is helpful that clearly distinguishes the problem(s), for which an AI is designed, and allows to specify the decision context along the dimensions as laid out in Table 2, and explicitly considers trade-offs, human decision behaviour and preferences. This approach emphasizes that no two crises or cities will ever be identical. Rather, they differ problem context, social networks, and actors (with their cognitive abilities and resources), and the guiding values, norms, and principles.

Discussion & conclusions

AI systems have tremendously influenced our lives. As our cities will be confronted with an increasing number and severity of crises, decision-makers are confronted with highly dynamic, volatile, and interacting processes and phenomena. Because of this complexity, especially in urban crisis response, there is widespread hope that AI can help us deal with the combination of complexity and urgency that is critical in crisis management. At the same time, this paper argues that AI introduces new dependencies and vulnerability, and may deepen or shift the crises that it is designed to resolve. I also argued that the insights from the area of crisis management in increasingly smart cities can be used for other contexts of 'super-wicked problems', ranging from adapting to a changing climate to pandemics, migration or poverty.

To guide the design and use of AI in society, various high-level AI guidances put forward a series of principles that AI should adhere to (for a review see (Jobin et al., 2019). Similarly, in the context of human-centred AI (Ozmen Garibay et al., 2023; Shneiderman, 2020), principles are set out to define the design process. Yet, in both cases, challenges persist in translating the principles to concrete design requirements or implementation for the purpose of urban crises. This paper has shown that crises decisions combine the well-known urgency and complexity with values, leading to a combination of cognitive and moral overload that AI is not yet adequately equipped to handel. First, while there is a lot of work on principles such as transparency or explainability, the fundamental principles that guide information management and decision-making in crisis response – such as humanity or solidarity – have not yet been formalized into machine-readable formats. Second, the urgency and high stakes of crises fundamentally change human information processing, sharing and decision-making – and therefore needs to be considered when designing for an AI that advises or interacts with humans in crises. Third, traditional models of automation need to be reconsidered that largely study one human interacting with one machine (Endsley, 2017; Parasuraman et al., 2000). In a crisis, there is no single decision-maker or 'operator'. Especially in densely populated urban areas, many different citizens, volunteers, emergency management professionals, businesses, and policy makers all act and decide. Decisions are therefore often informal or improvised, and far from unified, bringing about thousands of independent actions, and attempts to coordinate that may overlap, compete, interfere, cooperate, or even battle for the scarce resources available (Holguín-Veras et al., 2012). To be effective, AI systems have to be adaptive to the different constellations of people, machines resources, and information that come together in an urban crisis, and operate in a coordinated, yet flexible fashion (Jennings et al., 2014).

Throughout the paper, several questions have arisen, and below, I am briefly characterising the most urgent research needs for the responsible design of AI for crisis decision-making. As such, the paper aims to inspire decision-makers, researchers and AI developers and aims to make headway in understand how AI can be designed for the shifting timescapes that characterise crises and other super-wicked urban problems.

Designing for Human Control in volatile networks The urgency of crisis response, and the fear of human bias have led to a push for automation in crisis response. Yet, the volatile nature of crises often calls for improvisation and creativity, and machines need to be designed to be flexible adaptive to the changing contexts, and the emergent preferences and objectives. A critical issue is the balance of *agency* between the humans and the AI systems involved, especially under the time pressure of a crisis, when given the cognitive and moral overload, it may not be possible for humans to control or potentially override each AI system they interact with.

Even though there is a push to embed morality into machines (Arkin et al., 2012), acknowledging the context, the concepts of *adjustable autonomy* (Mostafa et al., 2019) and human *moral* autonomy, which defines the conditions that need to be fulfilled for human decision-makers to be able to maintain moral agency when interacting with an AI. Crisis AI needs to be designed to respect the cognitive strain on decision-makers while maintaining their moral autonomy, and they need to be dynamic, i.e., they need to 'learn' and adjust to the volatile roles and responsibilities as well as the changing situation at hand (Mendonca et al., 2006; Wallace & De Balogh, 1985). Besides theoretical work on defining how autonomy can be adjusted and upheld, interaction models and empirical research is needed on how humans (inter-)act with AI in crises. Given the possible pitfalls of experimentation in crises that were mentioned before, simulation games are a promising research methodology here, since they combine a safe space for trying and testing technology with relatively high fidelity and realism (Lukosch & Comes, 2019). To scale the insights from one or several humans to a whole city, agent-based models have been shown to help formalize and analyse the dynamic emergence of patterns (Helbing, 2012).

Designing AI for principles + urgency While remains a debate in crisis management if response should be guided by a rights based approach that focuses on capacities (Gready, 2008; Nussbaum, 2007), or by standards of care and care ethics (Leider et al., 2017; van Wynsberghe & Comes, 2020), there are several principles that have been suggested to guide crisis response, including solidarity and humanity. While there has been much work on areas such as explainability, accountability or fairness in AI that is translated into guidance, standards and recommendations (e.g., (EC, 2019; IEEE, 2019; OECD, 2019; UNESCO, 2022), these high-level principles are thus far neglected by the technology that mostly focuses on control and optimisation. An open question is how crisis-related principles relate to other principles (such as trustworthiness), and how they can be translated into protocols for data acquisition, or into an algorithm for information analysis or decision-making.

For the well-established principles of trustworthy or human-centred AI (Riedl, 2019), such as explainability or accountability, their definitions do not account for the emergent and networked nature of social-technical systems in crisis response. To ensure for instance explainability in urban crises, the high level principles will need to be refined to also take into account trade-offs with explanations vs. timeliness. By using a decision theoretical approach to explore trade-offs and preferences, crucial questions can be answered that lead to more concrete requiements, such as: what needs to be explainable and to whom in the heat of a crisis, or how to define accountability if negative effects arise from the interaction of many machines with many humans. This step requires a combination of empirical and theoretical work, ensuring that the behavioural insights on how humans do behave can be linked to the theoretical understanding of how they *should* behave.

Designing for the most vulnerable Crises affect all of society. Therefore, decisions around AI in crisis management are necessarily social choices. Yet, in crises, often those that are already vulnerable or marginalized are affected most. How can we ensure that AI is sensitive to their needs? Here, social choice ethics for AI design (Baum, 2020) advocates to consider questions around standing (who is heard?), measurement (what are the objectives, and how are they translated to data?), and aggregation (how do we form a choice based on potentially diverging preferences and attitudes?). This paper stressed that implementing ethical codes, values, norms and preferences of many stakeholders within AI systems is challenging (Sun & Medaglia, 2019). A major obstacle herein is creating a consensus across different viewpoints, given the difference in preferences and the potential power dynamics (Lewis et al., 2020). To ensure that there is true engagement, human-centred design approaches are needed that are fully inclusive. Besides the need to avoid biases in the way that damages are detected or resources are prioritized (Shams et al., 2023), this also means that vulnerable communities need to have a say about the fundamental principles of the technology that is designed to protect them. Here, methodologies are needed that help translate especially technical requirements into understandable language, and to ensure that people that may depend on assistance have the full autonomy to state their preferences.

Funding No funding was received to assist with the preparation of this manuscript.

Declarations

Competing interests The author has no competing interests to declare that are relevant to the content of this article.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Aitken, C., & Ersoy, E. (2023). War in Ukraine: The options for Europe's energy supply. *The World Economy*, *46*(4), 887–896.

Alshamaila, Y., Papagiannidis, S., Alsawalqah, H., & Aljarah, I. (2023). Effective use of smart cities in crisis cases: A systematic review of the literature. *International Journal of Disaster Risk Reduction*, *85*, 103521. https://doi.org/10.1016/j.ijdrr.2023.103521.

Arkin, R. C., Ulam, P., & Wagner, A. R. (2012). Moral decision making in autonomous systems: Enforcement, moral emotions, dignity, trust, and deception. *Proceedings of the IEEE*, *100*(3), 571–589. https://doi.org/10.1109/JPROC.2011.2173265.

Arndt, C., Diao, X., Dorosh, P., Pauw, K., & Thurlow, J. (2023). The Ukraine war and rising commodity prices: Implications for developing countries. *Global Food Security*, *36*, 100680.

Atif, M., Ahmad, R., Ahmad, W., Zhao, L., & Rodrigues, J. J. P. C. (2021). UAV-assisted wireless localization for search and rescue. *IEEE Systems Journal*, *15*(3), 3261–3272.

Atkinson, P., Gobat, N., Lant, S., Mableson, H., Pilbeam, C., Solomon, T., Tonkin-Crine, S., & Sheard, S. (2020). Understanding the policy dynamics of COVID-19 in the UK: Early findings from interviews with policy makers and health care professionals. *Social Science & Medicine*, *266*, 113423. https://doi.org/10.1016/j.socscimed.2020.113423.

Baharmand, H., Comes, T., & Lauras, M. (2019). Bi-objective multi-layer location-allocation model for the immediate aftermath of sudden-onset disasters. *Transportation Research Part E-Logistics and Transportation Review*, *127*, 86–110. https://doi.org/10.1016/j.tre.2019.05.002.

Batty, M. (2018). *Artificial intelligence and smart cities* (Vol. 45, pp. 3–6). SAGE Publications Sage UK.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, *214*(1), 481–518. https://doi.org/10.1140/epjst/e2012-01703-3.

Baum, S. D. (2020). Social choice ethics in artificial intelligence. *AI & Society*, *35*(1), 165–176. https://doi.org/10.1007/s00146-017-0760-1.

Bharosa, N., Lee, J., & Janssen, M. (2009). Challenges and obstacles in sharing and coordinating information during multi-agency disaster response: Propositions from field exercises. *Information Systems Frontiers*, *12*(1), 49–65. https://doi.org/10.1007/s10796-009-9174-z.

Bier, M., Fathi, R., Stephan, C., Kahl, A., Fiedrich, F., & Fekete, A. (2023). Spontaneous volunteers and the flood disaster 2021 in Germany: Development of social innovations in flood risk management. *Journal of Flood Risk Management*, *n/a*(n/a). https://doi.org/10.1111/jfr3.12933. https://doi.org/https://doi.org/10.1111/jfr3.12933.

Boin, A. (2019). The Transboundary Crisis: Why we are unprepared and the road ahead. *Journal of Contingencies and Crisis Management*, *27*(1), 94–99.

Boin, A., Ekengren, M., & Rhinard, M. (2016). The study of crisis management. *Routledge Handbook of Security Studies* (pp. 461–470). Routledge.

Bradshaw, J. M., Dignum, V., Jonker, C., & Sierhuis, M. (2012). Human-agent-robot teamwork. *IEEE Intelligent Systems*, *27*(2), 8–13.

Casali, Y., Yonca, N. A., Comes, T., & Casali, Y. (2022). Machine learning for spatial analyses in urban areas: A scoping review. *Sustainable Cities and Society*, 104050.

Champlin, C., Sirenko, M., & Comes, T. (2023). Measuring social resilience in cities: An exploratory spatio-temporal analysis of activity routines in urban spaces during Covid-19. *Cities*, *135*, 104220.

Coeckelbergh, M. (2020). Artificial Intelligence, responsibility attribution, and a relational justification of Explainability. *Science and Engineering Ethics*, *26*(4), 2051–2068. https://doi.org/10.1007/s11948-019-00146-8.

Comes, T., Hiete, M., Wijngaards, N., & Schultmann, F. (2011). Decision maps: A framework for multi-criteria decision support under severe uncertainty. *Decision Support Systems*, *52*(1), 108–118.

Comes, T., Van de Walle, B., & Van Wassenhove, L. (2020). The coordination-information bubble in humanitarian response: Theoretical foundations and empirical investigations. *Production and Operations Management*, *29*(11), 2484–2507.

Comes, T., Alexander, D., Boin, A., Eckert, C., Elmqvist, T., Fochesato, M., Helbing, D., Latusek-Jurczak, D., Lauta, K., & Meriläinen, E. (2022). Strategic crisis management in the European Union.

Comfort, L. K. (2007). Crisis management in hindsight: Cognition, communication, coordination, and control. *Public Administration Review*, *67*, 189–197.

Cools, J., Vanderkimpen, P., El Afandi, G., Abdelkhalek, A., Fockedey, S., Sammany, E., Abdallah, M., El Bihery, G., Bauwens, M., W., & Huygens, M. (2012). An early warning system for flash floods in hyper-arid Egypt. *Natural Hazards and Earth System Sciences*, *12*(2), 443–457.

Coppi, G., Jimenez, M., R., & Kyriazi, S. (2021). Explicability of humanitarian AI: A matter of principles. *Journal of International Humanitarian Action*, *6*, 1–22.

Council, N. R. (1998). In C. Wickens, A. Mavor, R. Parasuraman, & J. McGee (Eds.), *The future of air traffic control: Human operators and automation*. National Academies Press.

Crawford, K., & Finn, M. (2015). The limits of crisis data: Analytical and ethical challenges of using social and mobile data to understand disasters. *Geojournal*, *80*, 491–502.

de Bruijn, K. M., Jafino, B. A., Merz, B., Doorn, N., Priest, S. J., Dahm, R. J., Zevenbergen, C., Aerts, J. C., & Comes, T. (2022). Flood risk management through a resilience lens. *Communications Earth & Environment*, *3*(1), 285.

Deininger, K., Ali, D. A., Kussul, N., Shelestov, A., Lemoine, G., & Yailimova, H. (2023). Quantifying war-induced crop losses in Ukraine in near real time to strengthen local and global food security. *Food Policy*, *115*, 102418.

EC (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Elebe, T. M., & Kurnaz, S. (2023). Efficient detection of refugees and migrants in Turkey using convolutional neural network. *Physical Communication*, *59*, 102078.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*(1), 5–27.

Främling, K., & Explainable (2020). Transparent autonomous agents and multi-agent systems. Cham.

French, S., & Geldermann, J. (2005). The varied contexts of environmental decision problems and their implications for decision support. *Environmental Science & Policy*, *8*(4), 378–391.

Gao, H., Barbier, G., & Goolsby, R. (2011). Harnessing the Crowdsourcing Power of Social Media for Disaster Relief. *IEEE Intelligent Systems*, *26*(3), 10–14. https://doi.org/10.1109/MIS.2011.52.

Grasso, V. F., Beck, J. L., & Manfredi, G. (2007). Seismic early warning systems: Procedure for automated decision making. *Earthquake Early Warning Systems*, 179–209.

Gready, P. (2008). Rights-based approaches to development: What is the value-added? *Development in Practice*, *18*(6), 735–747.

Guikema, S. (2020). Artificial intelligence for natural hazards risk analysis: Potential, challenges, and research needs. *Risk Analysis*, *40*(6), 1117–1123. https://doi.org/10.1111/risa.13476.

Hanselmann, M., & Tanner, C. (2008). Taboos and conflicts in decision making: Sacred values, decision difficulty, and emotions. *Judgment and Decision Making*, *3*(1), 51–63.

Helbing, D. (2009). Managing complexity in socio-economic systems. *European Review*, *17*, 423–438.

Helbing, D. (2012). Agent-based modeling. *Social self-organization: Agent-based simulations and experiments to study emergent social behavior* (pp. 25–70). Springer.

Herath, H., & Mittal, M. (2022). Adoption of artificial intelligence in smart cities: A comprehensive review. *International Journal of Information Management Data Insights*, *2*(1), 100076.

Höchtl, J., Parycek, P., & Schöllhammer, R. (2016). Big data in the policy cycle: Policy decision making in the digital era. *Journal of Organizational Computing and Electronic Commerce*, *26*(1–2), 147–169. https://doi.org/10.1080/10919392.2015.1125187.

Holguin-Veras, J., Perez, N., Jaller, M., Van Wassenhove, L. N., & Aros-Vera, F. (2013). On the appropriate objective function for post-disaster humanitarian logistics models. *Journal of Operations Management*, *31*(5), 262–280. https://doi.org/10.1016/j.jom.2013.06.002.

Holguín-Veras, J., Jaller, M., Van Wassenhove, L. N., Pérez, N., & Wachtendorf, T. (2012). On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, *30*(7–8), 494–506. https://doi.org/10.1016/j.jom.2012.08.003.

IEEE. (2019). Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent systems. *Ethically Aligned Design - A Vision for Prioritizing Human Well-being with Autonomous and Intelligent systems* (pp. 1–294). IEEE.

Ishmaev, G., Dennis, M., & van den Hoven, M. J. (2021). Ethics in the COVID-19 pandemic: Myths, false dilemmas, and moral overload. *Ethics and Information Technology*, *23*(1), 19–34. https://doi.org/10.1007/s10676-020-09568-6.

Jennings, N. R., Moreau, L., Nicholson, D., Ramchurn, S., Roberts, S., Rodden, T., & Rogers, A. (2014). Human-agent collectives. *Communications of the ACM*, *57*(12), 80–88.

Jobin, A., Ienca, M., & Vayena, E. (2019). The global landscape of AI ethics guidelines. *Nature Machine Intelligence*, *1*(9), 389–399. https://doi.org/10.1038/s42256-019-0088-2.

Jørgensen, F., Bor, A., Rasmussen, M. S., Lindholt, M. F., & Petersen, M. B. (2022). Pandemic fatigue fueled political discontent during the COVID-19 pandemic. *Proceedings of the National Academy of Sciences*, *119*(48), e2201266119.

Josephson, A., Kilic, T., & Michler, J. D. (2021). Socioeconomic impacts of COVID-19 in low-income countries. *Nature Human Behaviour*, *5*(5), 557–565. https://doi.org/10.1038/s41562-021-01096-7.

Kankanamge, N., Yigitcanlar, T., Goonetilleke, A., & Kamruzzaman, M. (2020). Determining disaster severity through social media analysis: Testing the methodology with South East Queensland Flood tweets. *International Journal of Disaster Risk Reduction*, *42*, 101360.

Keeney, R. L. (1996). Value-focused thinking: Identifying decision opportunities and creating alternatives. *European Journal of Operational Research*, *92*(3), 537–549.

Kitchin, R., & Dodge, M. (2020). The (in) security of smart cities: Vulnerabilities, risks, mitigation, and prevention. *Smart cities and innovative Urban technologies* (pp. 47–65). Routledge.

Klein, G., Calderwood, R., & Clinton-Cirocco, A. (2010). Rapid decision making on the Fire Ground: The original study plus a Postscript. *Journal of Cognitive Engineering and Decision Making*, *4*(3), 186–209. https://doi.org/10.1518/155534310X12844000801203.

Kleiner, A., & Dornhege, C. (2007). Real-time localization and elevation mapping within urban search and rescue scenarios. *Journal of Field Robotics*, *24*(8-9), 723–745.

Kong, L., & Woods, O. (2018). The ideological alignment of smart urbanism in Singapore: Critical reflections on a political paradox. *Urban Studies*, *55*(4), 679–701. https://doi.org/10.1177/0042098017746528.

Krafft, P. M., Young, M., Katell, M., Huang, K., & Bugingo, G. (2020). 2020). Defining AI in policy versus practice.

Krishnan, S., Aydin, N. Y., & Comes, M. (2024). TIMEWISE: Temporal dynamics for Urban Resilience: Theoretical insights and empirical reflections from Amsterdam and Mumbai. *npj Urban Sustainability, 4*(1), 4.

Kumar, A., & Parikh, S. (2023). 12–13 May 2023). ML Based Automated Assistance System for Efficient Crowd Control A detailed investigation. 2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE).

Leider, J. P., DeBruin, D., Reynolds, N., Koch, A., & Seaberg, J. (2017). Ethical guidance for disaster response, specifically around crisis standards of care: A systematic review. *American Journal of Public Health*, *107*(9), e1–e9.

Lentz, E. C., Michelson, H., Baylis, K., & Zhou, Y. (2019). A data-driven approach improves food insecurity crisis prediction. *World Development*, *122*, 399–409.

Levin, K., Cashore, B., Bernstein, S., & Auld, G. (2012). Overcoming the tragedy of super wicked problems: Constraining our future selves to ameliorate global climate change. *Policy Sciences*, *45*(2), 123–152.

Lewis, D., Hogan, L., Filip, D., & Wall, P. J. (2020). Global challenges in the standardization of ethics for trustworthy AI. *Journal of ICT Standardization*, 123–150.

Lukosch, H., & Comes, T. (2019). Gaming as a research method in humanitarian logistics. *Journal of Humanitarian Logistics and Supply Chain Management*, *9*(3), 352–370. https://doi.org/10.1108/JHLSCM-06-2018-0046.

Mahajan, S., Hausladen, C. I., Sánchez-Vaquerizo, J. A., Korecki, M., & Helbing, D. (2022). Participatory resilience: Surviving, recovering and improving together. *Sustainable Cities and Society*, *83*, 103942.

Martini, G., Bracci, A., Riches, L., Jaiswal, S., Corea, M., Rivers, J., Husain, A., & Omodei, E. (2022). Machine learning can guide food security efforts when primary data are not available. *Nature Food*, *3*(9), 716–728.

Mendonca, D., Beroggi, G. E. G., van Gent, D., & Wallace, W. A. (2006). Designing gaming simulations for the assessment of group decision support systems in emergency response. *Safety Science*, *44*(6), 523–535.

Mendonça, D., Jefferson, T., & Harrald, J. (2007). Collaborative adhocracies and mix-and-match technologies in emergency management. *Communications of the ACM*, *50*, 44–49.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, *267*, 1–38. https://doi.org/10.1016/j.artint.2018.07.007.

Mittelstadt, B. (2019). Principles alone cannot guarantee ethical AI. *Nature Machine Intelligence*, *1*(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4.

Mostafa, S. A., Ahmad, M. S., & Mustapha, A. (2019). Adjustable autonomy: A systematic literature review. *Artificial Intelligence Review*, *51*(2), 149–186. https://doi.org/10.1007/s10462-017-9560-8.

Munawar, H. S., Mojtahedi, M., Hammad, A. W. A., Kouzani, A., & Mahmud, M. A. P. (2022). Disruptive technologies as a solution for disaster risk management: A review. *Science of the Total Environment*, *806*, 151351. https://doi.org/10.1016/j.scitotenv.2021.151351.

Nespeca, V., Comes, T., Meesters, K., & Brazier, F. (2020). Towards coordinated self-organization: An actor-centered framework for the design of disaster management information systems. *International Journal of Disaster Risk Reduction*, *51*, 101887. https://doi.org/10.1016/j.ijdrr.2020.101887.

Nussbaum, M. (2007). Human rights and human capabilities. *Harv Hum Rts J*, *20*, 21.

OECD (2019). *Recommendation of the Council on Artificial Intelligence*. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

Ogie, R. I., Rho, J. C., & Clarke, R. J. (2018). 4–7 Dec. 2018). Artificial Intelligence in Disaster Risk Communication: A Systematic Literature Review. 2018 5th International Conference on Information and Communication Technologies for Disaster Management (ICT-DM).

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C., Falco, G., Fiore, S. M., Garibay, I., Grieman, K., Havens, J. C., Jirotka, M., Kacorri, H., Karwowski, W., Kider, J., Konstan, J., Koon, S., Lopez-Gonzalez, M., Maifeld-Carucci, I., & Xu, W. (2023). Six human-centered Artificial Intelligence Grand challenges. *International Journal of Human–Computer Interaction*, *39*(3), 391–437. https://doi.org/10.1080/10447318.2022.2153320.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems Man and Cybernetics - Part A: Systems and Humans*, *30*(3), 286–297. https://doi.org/10.1109/3468.844354.

Paulus, D., Fathi, R., Fiedrich, F., de Walle, B. V., & Comes, T. (2022). On the Interplay of Data and Cognitive Bias in Crisis Information Management. *Information Systems Frontiers*. https://doi.org/10.1007/s10796-022-10241-0.

Paulus, D., de Vries, G., Janssen, M., & Van de Walle, B. (2023). Reinforcing data bias in crisis information management: The case of the Yemen humanitarian response. *International Journal of Information Management*, *72*, 102663. https://doi.org/10.1016/j.ijinfomgt.2023.102663.

Pauwels, N., Van De Walle, B., Hardeman, F., & Soudan, K. (2000). The implications of irreversibility in emergency response decisions. *Theory and Decision*, *49*(1), 25–51.

Persson, E., & Tinghög, G. (2023). The effect of fast and slow decision-making on equity–efficiency tradeoffs and moral repugnance. *Royal Society Open Science*, *10*(9), 230558. https://doi.org/10.1098/rsos.230558.

Prainsack, B., Neves, M. C. P., Sahlin, N. E., Biller-Andorno, N., Laukyte, M., Łuków, P., Nys, H., van den Hoven, J., Mallia, P., & Molnár-Gábor, F. (2023). Values in challenging times: strategic crisis management in the EU. *The Lancet Regional Health–Europe*, *24*.

Qadir, J., Ali, A., Zwitter, A., Sathiaseelan, A., & Crowcroft, J. (2016). Crisis analytics: Big data-driven crisis response. *Journal of International Humanitarian Action*, *1*(1), 1–21.

Quarantelli, E. L. (1988). Disaster Crisis Management: A Summary of Resarch findings. *Journal of Management Studies*, *25*(4), 373–385. https://doi.org/10.1111/j.1467-6486.1988.tb00043.x.

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J. F., Breazeal, C., Crandall, J. W., Christakis, N. A., Couzin, I. D., & Jackson, M. O. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486.

Renn, O., & Lucas, K. (2021). Systemic risk: The threat to Societal Diversity and Coherence. *Risk Analysis*, *n/a*(n/a). https://doi.org/10.1111/risa.13654. https://doi.org/https://doi.org/10.1111/risa.13654.

Riedl, M. O. (2019). Human-centered artificial intelligence and machine learning. *Human Behavior and Emerging Technologies*, *1*(1), 33–36.

Rizk, M., Slim, F., & Charara, J. (2021). 2021). Toward AI-assisted UAV for human detection in search and rescue missions.

Rosenthal, U., Charles, M. T., & Hart, P. (1989). *Coping with crises: The management of disasters, riots, and terrorism*. Charles C Thomas Pub Limited.

Sandvik, K. B., Jacobsen, K. L., & McDonald, S. M. (2017). Do no harm: A taxonomy of the challenges of humanitarian experimentation. *International Review of the Red Cross*, *99*(904), 319–344. https://doi.org/10.1017/S181638311700042X.

Seneviratne, S. I., Zhang, X., Adnan, M., Badi, W., Dereczynski, C., Di Luca, A., Vicente-Serrano, S. M., Wehner, M., & Zhou, B. (2021). 11 Chap. 11: weather and climate extreme events in a changing climate.

Shams, R. A., Zowghi, D., & Bano, M. (2023). AI and the quest for diversity and inclusion: A systematic literature review. *AI and Ethics*, 1–28.

Sheridan, T. B., Verplank, W. L., & Brooks, T. L. (1978). 1978). Human/computer control of undersea teleoperators.

Shneiderman, B. (2020). Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Transactions on Interactive Intelligent Systems*, *10*(4). https://doi.org/10.1145/3419764.

Sigala, I. F., Sirenko, M., Comes, T., & Kovács, G. (2022). Mitigating personal protective equipment (PPE) supply chain disruptions in pandemics–a system dynamics approach. *International Journal of Operations & Production Management*, *42*(13), 128–154.

Sloane, M., Solano-Kamaiko, I. R., Yuan, J., Dasgupta, A., & Stoyanovich, J. (2023). Introducing contextual transparency for automated decision systems. *Nature Machine Intelligence*, *5*(3), 187–195. https://doi.org/10.1038/s42256-023-00623-7.

Strauch, B. (2017). Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems*, *48*(5), 419–433.

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, *36*(2), 368–383. https://doi.org/10.1016/j.giq.2018.09.008.

Sun, W., Bocchini, P., & Davison, B. D. (2020). Applications of artificial intelligence for disaster management. *Natural Hazards*, *103*(3), 2631–2689.

Szocik, K., & Abylkasymova, R. (2022). Ethical issues in Police Robots. The case of Crowd Control Robots in a pandemic. *Journal of Applied Security Research*, *17*(4), 530–545. https://doi.org/10.1080/19361610.2021.1923365.

Tschunkert, K., & Vogel, B. (2023). Humanitarian cash assistance: Risks and opportunities for local peace. *Development in Practice*, *33*(5), 560–571. https://doi.org/10.1080/09614524.2022.2134302.

Tsoukias, A., Montibeller, G., Lucertini, G., & Belton, V. (2013). Policy analytics: An agenda for research and practice. *EURO Journal on Decision Processes*, *1*(1), 115–134. https://doi.org/10.1007/s40070-013-0008-3.

Tsoukiàs, A. (2008). From decision theory to decision aiding methodology. *European Journal of Operational Research*, *187*(1), 138–161. https://doi.org/10.1016/j.ejor.2007.02.039.

Turoff, M., Chumer, M., Van de Walle, B. A., & Yao, X. (2004). The design of a dynamic emergency response management information system. *Journal of Information Technology Theory and Applications*, *5*(4), 1–36. http://www.narcis.nl/publication/RecordID/oai:wo.uvt.nl:180223.

Twigg, J., & Mosel, I. (2017). Emergent groups and spontaneous volunteers in urban disaster response. *Environment and Urbanization*, *29*(2), 443–458.

Tzachor, A., Whittlestone, J., Sundaram, L., & hÉigeartaigh, S. (2020). Artificial intelligence in a crisis needs ethics with urgency. *Nature Machine Intelligence*, *2*(7), 365–366. https://doi.org/10.1038/s42256-020-0195-0.

UN-OCHA (2016). *Building data responsibility into humanitarian action* (OCHA Policies and Studies Series, Issue. https://www.unocha.org/sites/unocha/files/Building%20data%20responsibility%20into%20humanitarian%20action.pdf.

UNESCO (2022). *Recommendation on the Ethics of Artificial Intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000381137.

Ussher, L., Ebert, L., Gómez, G. M., & Ruddick, W. O. (2021). Complementary currencies for humanitarian aid. *Journal of Risk and Financial Management*, *14*(11), 557.

Van de Walle, B., & Comes, T. (2015). On the nature of information management in complex and natural disasters. *Procedia Engineering*, *107*, 403–411.

Van de Walle, B., Brugghemans, B., & Comes, T. (2016). Improving situation awareness in crisis response teams: An experimental analysis of enriched information and centralized coordination. *International Journal of Human-Computer Studies*. https://doi.org/10.1016/j.ijhcs.2016.05.001.

Van Wynsberghe, A. (2021). Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, *1*(3), 213–218.

van Wynsberghe, A., & Comes, T. (2020). Drones in humanitarian contexts, robot ethics, and the human–robot interaction. *Ethics and Information Technology*, *22*(1), 43–53.

Voigt, S., Kemper, T., Riedlinger, T., Kiefl, R., Scholte, K., & Mehl, H. (2007). Satellite image analysis for disaster and crisis-management support. *IEEE Transactions on Geoscience and Remote Sensing*, *45*(6), 1520–1528.

Wallace, W. A., & De Balogh, F. (1985). Decision support systems for disaster management. *Public Administration Review*, 134–146.

Wang, S., & Li, M. (2021). 12–14 Nov. 2021). Research on public safety emergency management of Smart city. 2021 2nd International Conference on Computer Science and Management Technology (ICCSMT).

WEF (2023). *The Global Risks Report 2023*. https://www3.weforum.org/docs/WEF_Global_Risks_Report_2023.pdf.

Weick, K. E., Sutcliffe, K. M., & Obstfeld, D. (2005). Organizing and the process of Sensemaking. *Organization Science*, *16*(4), 409–421. https://doi.org/10.1287/orsc.1050.0133.

Wilkinson, E., Weingärtner, L., Choularton, R., Bailey, M., Todd, M., Kniveton, D., & Venton, C. (2018). C. *Forecasting hazards, averting disasters: implementing forecast-based early action at scale*.

Wolbers, J., Boersma, K., & Groenewegen, P. (2018). Introducing a fragmentation perspective on coordination in crisis management. *Organization Studies*, *39*(11), 1521–1546.

Yang, C., Su, G., & Chen, J. (2017). 2017). Using big data to enhance crisis response and disaster resilience for a smart city.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.