Scientific Advice Mechanism
to the European Commission

December 2025

SAPEA Rapid Evidence Review Report

Artificial Intelligence in Emergency and Crisis Management

Funded by the European Union. Views and opinions expressed are however those of SAPEA, working group members and other experts involved only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor the European Commission can be held responsible for them.

- SAPEA (2025). *Artificial intelligence in emergency and crisis management: Rapid evidence review report*. Munich: SAPEA.
- 10.5281/zenodo.17737962
- Downloadable from https://doi.org/10.5281/zenodo.17737962

Version history

Version	Date	Summary of changes
1.0	11 December 2025	Publication

Scientific Advice Mechanism
to the European Commission

Artificial Intelligence in Emergency and Crisis Management

SAPEA Rapid Evidence Review Report

December 2025

Table of contents

List of abbreviations

Abbreviation	Full form
AABA	AI Is Better At
ABT	Augmented Bird Table
AI	Artificial Intelligence
API	Application Programming Interface
BERT	Bidirectional Encoder Representations from Transformer (Google model)
CAP	Common Alerting Protocol
CBRN	Chemical, Biological, Radiological, Nuclear
CDC	Centers for Disease Control and Prevention
COD	Common Operational Data
CV	Computer Vision
DAS	Data Structures and Algorithms
DG ECHO	Directorate-General for European Civil Protection and Humanitarian Aid Operations
DL	Deep Learning
DPIA	Data Protection Impact Assessment
DRMKC	Disaster Risk Management Knowledge Centre
ECDC	European Centre for Disease Prevention and Control
ERCC	Emergency Response Coordination Centre
EWS	Early Warning Systems
FAIR	Findable, Accessible, Interoperable and Reusable
FEWS NET	Famine Early Warning Systems Network
GDACS	Global Disaster Alert and Coordination System
GDPR	General Data Protection Regulation
GPAI	General Purpose AI
GPS	Global Positioning System
GPT	Generative Pre-Trained Transformer
HABA	Humans Are Better At
HCAI	Human-Centred AI
HCI	Human-Computer Interaction
ICRC	International Committee of the Red Cross
IFRC	International Federation of Red Cross and Red Crescent Societies
INSPIRE	Information for Spatial Infrastructure in Europe
IoT	Internet of Things
LLMs	Large Language Models
MABA	Machines Are Better At
ML	Machine Learning
NGO	Non-governmental organisation
NLP	Natural Language Processing
NOAA	National Oceanic and Atmospheric Administration
ODD	Operating Decision Domain

OECD	Organisation for Economic Cooperation and Development
RAG	Retrieval-Augmented Generation
SAM	(European) Scientific Advice Mechanism
Tax-CIM	Taxonomy for Crisis Information Management Systems
UAV	Unmanned Aerial Vehicles
UCPM	Union Civil Protection Mechanism
UNDRR	United Nations Office for Disaster Risk Reduction
UNESCO	United Nations Educational, Scientific and Cultural Organisation
UN-OCHA	United Nations Office for the Coordination of Humanitarian Affairs
WHO	World Health Organization
XAI	Explainable AI

Foreword

Over the past decade, the development and deployment of Artificial Intelligence have accelerated significantly. What was once confined largely to some research and industry sectors has now entered almost every aspect of our lives, thus becoming a societal, economic and political priority. Important debates and questions accompany the growing use of AI. One of those most pressing questions is how we can benefit from the full potential of these technologies, while also understanding and managing the risks that come with them.

In 2022, the Scientific Advice Mechanism delivered advice on how to improve Strategic Crisis Management in the European Union, in response to a request from the European Commission's Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO). This has stimulated further reflection and become the starting point for a new request for advice on crisis management by the Emergency Response Coordination Centre (ERCC) within DG ECHO. The focus this time is on the characteristics, opportunities and risks associated with the use of AI in crisis preparedness and response. Recent crises, including the COVID-19 pandemic, extreme weather events, and geopolitical tensions, have underscored the urgency of strengthening Europe's crisis preparedness and response capacities, making the consideration of AI particularly timely. AI systems, however, must be used responsibly, ethically, and transparently in order to maintain public trust and safeguard fundamental rights.

This is a *Rapid Evidence Review Report*. To address the highly complex nature of the topic in the requested timeframe, SAPEA assembled a small interdisciplinary working group of outstanding experts in the field and then involved a wider group of experts to review the Report.

The project was coordinated by ALLEA, the European Federation of Academies of Sciences and Humanities, acting as the lead network on behalf of SAPEA. Cardiff University acted as collaborating institution.

We warmly thank all contributing experts for their time and contributions, in addition to everyone else involved in assembling this Report. We wish to express our particular appreciation to the Working Group Chair, Professor Tina Comes, who has shown boundless dedication, energy and enthusiasm in this role.

We would also like to express our sincere gratitude to the science academies across Europe, thanks to whom SAPEA can bring together the best available science.

Professor **Paweł Rowiński,** President of ALLEA
Professor **Donald Dingwell,** Chair of the SAPEA board

Preface

Heatwaves, floods, droughts and storms, the repercussions of the COVID-19 pandemic and the war in Ukraine have all affected the lives and livelihoods of millions of Europeans. Challenges like these can unfold simultaneously across multiple sectors, creating cascading effects that demand new approaches to preparedness and response.

Artificial intelligence (AI) can improve crisis preparedness, response and recovery by collecting, processing, and analysing unprecedented amounts of data, at rapid speed. At the same time, there are important legal, ethical and institutional concerns to be addressed, ensuring that AI is used in a trustworthy and responsible way.

We have developed this Rapid Evidence Review Report in response to a request to the Scientific Advice Mechanism (SAM) from the European Commission's Emergency Response Coordination Centre (ERCC) within the Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO). A small interdisciplinary Working Group drafted this evidence brief over a concentrated timeframe. Rather than the usual period of around 9 to 12 months for a full Evidence Review Report, this Rapid Evidence Review Report has been completed in less than 6 months.

The approach has allowed us to respond to an urgent policy need, but it has also shaped our methodology significantly. We have conducted a review that is targeted, incorporating evidence from peer-reviewed sources and grey literature. We have prioritised recent publications on AI, while recognising that foundational work on crisis management remains relevant.

Where appropriate, we draw on other fields, such as health or generic AI literature. However, we note that crises come with specific constraints, such as urgency, time pressure and complexity, which distinguish crisis management from other domains. While for some challenges, such as weather prediction, there exists an objective 'ground truth' (direct empirical evidence) against which AI performance can be compared, it becomes more complicated for questions involving human behaviour, vulnerability, or resilience, where the phenomenon itself is context dependent. This distinguishes it from fields such as evidence-based medicine, for example, which has a long tradition of methods like randomised controlled trials (RCTs).

The Report makes several deliberate positioning choices that reflect the opportunities and the challenges the Working Group encountered:

AI is more than ChatGPT. Even though the current policy debate focuses very much on large language models, there is promise in different types of AI systems and applications. Consequently, we first take a close look at AI as an 'umbrella' of different methods, approaches and tools, providing examples and case studies to illustrate the different areas. These range from machine learning for extreme weather forecasting, to large language models for tracking and countering misinformation.

Principles over tools. Given that AI research and practice are developing so rapidly, we did not attempt to review different tools that are currently available on the market, since such inventories would be rapidly outdated. Instead, we provide the reader with guiding principles on *how to choose an AI tool* for different phases of crisis management, alongside legal and ethical considerations that should inform their selection and use. This approach will enable users to assess new or updated AI tools, as they become available. Similarly, the reliability and performance of AI tools are sensitive to the specific hazard, its geographic setting, data availability, institutional arrangements, and user capabilities. A system that performs well for flood forecasting in one river basin may fail in another due to differences in hydrology, sensor coverage, or historical data. Rather than providing potentially misleading universal maturity scores, we offer criteria and boundary conditions that allow users to evaluate whether a particular AI system is fit for purpose in the specific context.

AI must support humans. We view AI in crisis management not merely as a computational tool; rather, we examine it through a socio-technical systems lens. This means we recognise that AI influences the way people collect and share information, make sense of their environment, and eventually make decisions. Given the goal of AI is to enhance human capabilities rather than replacing human judgement, we integrate literature relating to Human-Centred AI and Hybrid Intelligence.

Crises do not respect national boundaries. AI has to reflect the reality of cross-border operations and multi-national crisis response. This requires data sharing between Member States, the interoperability of AI systems, and coordination mechanisms that respect EU legislation (the AI Act and GDPR). Using AI for European crisis response therefore depends on institutional capacity-building, training programmes that span national civil protection authorities, and data preparedness frameworks that enable rapid data sharing while protecting privacy.

In conclusion, sustainable progress in AI for crisis management depends on building institutional capacity to evaluate critically, govern responsibly, and use AI systems appropriately within Europe's crisis management architecture. The frameworks and principles we present aim to support evidence-based decision-making, as Europe navigates the opportunities and challenges of AI in an evolving threat landscape. The Report closes with a catalogue of policy options, along with advantages and challenges designed to foster a discussion on how to strengthen European Crisis Management AI capabilities.

The Working Group on AI and Crisis Management

Working Group Members

- **Tina Comes** (Chair), German Aerospace Center & TU Delft
- **Verónica Bolón-Canedo**, Universidade da Coruña
- **Joachim Denzler**, Friedrich Schiller University Jena
- **Nick Jennings**, Loughborough University
- **Thomas Kox**, Weizenbaum Institute
- **Markus Reichstein**, Max-Planck Institute for Biogeochemistry & Friedrich Schiller University Jena
- **Christian Reuter,** TU Darmstadt
- **Andrej Zwitter**, University of Groningen

Contributor: **Olya Kudina**, TU Delft

Executive Summary

This *SAPEA Rapid Evidence Review Report* has been produced in response to a request to the Scientific Advice Mechanism (SAM) from the European Commission's Emergency Response Coordination Centre (ERCC) within the Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO).

The ERCC asked the SAM to synthesise current knowledge on the use of AI in emergency and crisis management, identifying opportunities and risks, and how these can be mitigated.

The Report has been drafted by a small Working Group of independent experts. It addresses:

Definitions, framing and scope. AI in emergency and crisis management covers a range of technologies and methods that support or automate tasks that would typically involve human input and intelligence. Types of AI include machine learning, natural language processing, computer vision, simulations, generative AI and agentic AI. AI can help with situational awareness, forecasting, damage assessment; it can also provide decision support across the disaster risk management cycle of prevention, preparedness, response, and recovery. At the same time, the use of AI must uphold human dignity, transparency, and responsibility, as well as meeting international standards of safety, ethics, and crisis governance.

Requirements of AI tools. AI tools in crisis contexts must be usable by a range of practitioners in the field. The Report sets out a classification that outlines the required functionalities of AI tools, providing guidance in assessing, implementing, using and regulating such tools in crisis contexts.

Performance of AI. The usefulness of AI depends on whether its capabilities fit the tasks at hand, and whether it can support existing organisational processes and protocols. This Report examines the performance of AI in three areas of crisis management: monitoring and predicting, assessing, and decision-making. It also identifies areas where the use of AI is not appropriate or is problematic. Looking through a socio-technical lens, it highlights the evolution towards hybrid systems, where humans and AI cooperate in teams.

Legislative and ethical frameworks. The Report considers critical legislation, such as the EU's AI Act and the GDPR, together with 'soft law', such as ethical principles, frameworks and standards. It provides illustrative examples to show how European legal frameworks may apply in practice.

Data challenges and ways forward. The Report examines data sharing and governance, highlighting challenges such as data quality and availability, issues of privacy and legal compliance, cross-border and inter-agency coordination. It puts forward possible requirements for a new framework, with the aim of strengthening Europe's data preparedness.

User uptake and trust. The Report considers issues of user uptake of AI in crisis management. It highlights the importance of building trust in AI systems, which is crucial to uptake, and the need to avoid overconfidence in AI to ensure meaningful human control.

Case studies and examples of applications. The Report includes examples of applications and case studies, as a means to illustrate some of the key points. The four case studies cover disinformation detection, weather forecasting, disaster response in Nepal, and the COVID-19 pandemic.

Conclusions. The Report concludes that AI excels at standardised, data-intensive tasks that are typical in frequent disasters such as floods, wildfires or droughts. AI can be effective in environmental monitoring, early-warning systems, damage assessment from satellite imagery, and social media processing. The evidence suggests that AI performs better in well-defined crisis management tasks, particularly those that involve rapid processing and pattern recognition of large volumes of heterogeneous data. It is not yet well-suited to interpreting contexts that are highly varied, or for use in situations that are new or which lack suitable training data, or where moral choices and/or trade-offs are involved. Careful monitoring is also required to ensure compliance with legal and governance frameworks, avoid algorithmic biases and provide appropriate human control.

Policy options. *The Report presents a catalogue of possible policy options, based on the evidence. These include:*

- Establishing a European Data Preparedness Framework
- Integrating AI literacy into crisis training programmes
- Developing AI evaluation benchmarks and knowledge-sharing platforms
- Advancing European strategic autonomy for crisis AI
- Ensuring full compliance with legal frameworks
- Clarifying legal responsibilities in cross-border and public-private operations
- Addressing legal and ethical gaps in non-EU humanitarian operations
- Strengthening oversight, transparency and bias mitigation in AI tools
- Ensuring ethical AI through EU-level strategic governance, for example, on data and modelling.

1. Introduction

Artificial intelligence (AI) is rapidly changing the way that data is collected, analysed and processed, to then inform actions or decision-making. A recent survey in the *Stanford AI Index Report 2025* shows that nearly 80% of respondents say that their organisation uses AI for at least one function (up from 55% in 2023), with remarkable growth in the use of generative AI[1]. Against this backdrop, many active discussions are taking place, at both national and European levels, on how AI can be used in emergency and crisis management. With recent advances and the variety of AI tools available, there is an identified need to synthesise the evidence on the performance of these tools across a range of crisis management tasks; to examine the existing frameworks for assessing AI capabilities, based on their application; and to outline lessons learned about the current reliability and maturity of such applications, in view of experience from real-world implementation.

The European Commission's Emergency Response Coordination Centre (ERCC) within the Directorate-General for European Civil Protection and Humanitarian Aid Operations (DG ECHO) has requested the European Scientific Advice Mechanism (SAM) to deliver a Rapid Evidence Review Report that consolidates current knowledge on AI applications in crisis management and provides frameworks for understanding their capabilities and limitations, based on the existing literature. The Report may also inform the future integration and use of AI in emergency and crisis management, both for the Emergency Response Coordination Centre (ERCC) and, more generally, for crises centres across Europe.

The primary questions posed are as follows:

Based on the evidence, what are the characteristics, opportunities and risks associated with the use of artificial intelligence in crisis preparedness and response? According to the literature, how can these risks be mitigated?

A small interdisciplinary Working Group, composed of independent experts, has drafted this concise evidence-based report to address DG ECHO's request. It first sets out definitions and considers the framing of the topic. It then examines AI's performance across a range of tasks, identifying challenges and potential ways forward. The Report includes several case studies and examples, which demonstrate how AI has been implemented in a range of real-world settings. Lastly, the Report sets out its conclusions and puts forward a number of options for policy, based on the evidence. This Report is complemented by an introductory narrative review of recent literature, which is published alongside it.

As stated in the Preface, the Report describes principles and frameworks, rather than assessing specific AI tools. This approach is intended to enable the reader of the Report to assess the potential of AI tools, whether commercial or open access, for different areas of application and across a range of institutional settings.

1 https://hai.stanford.edu/assets/files/hai_ai_index_report_2025.pdf, p. 261

2. Definitions and scope

Introduction

Artificial Intelligence (AI) in emergency and crisis management refers to computational systems designed to support or automate data collection, information processing, forecasting and decision-making throughout the disaster risk management[2] cycle. This includes prevention and preparedness (risk assessment, forecasting, public awareness tools), response (real-time decision support, resource allocation, communication), and recovery (damage assessment, coordination of rebuilding, long-term impact modelling). AI systems operate at the intersection between rapid, complex decision demands and heterogeneous data environments (Comes, 2024).

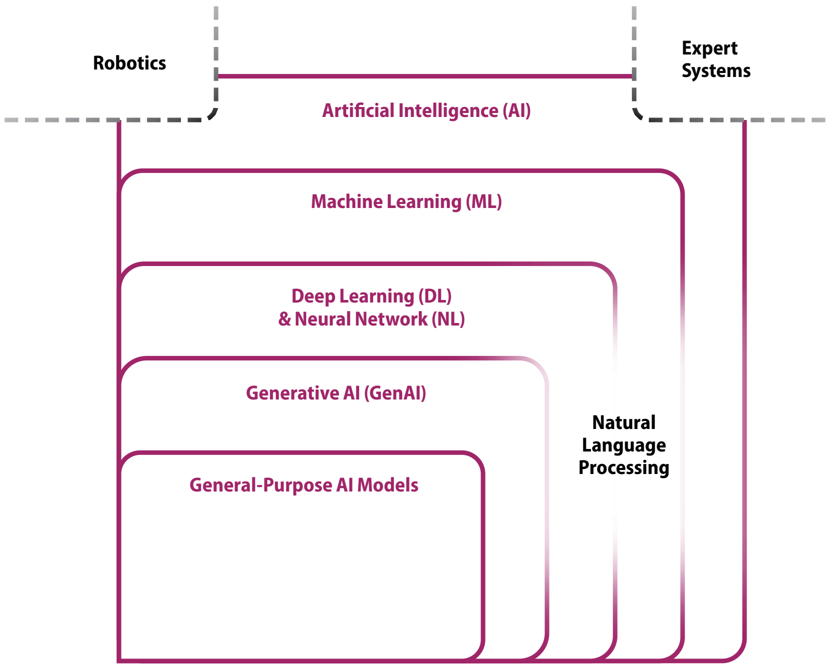


Figure 1. What is Artificial Intelligence? An overview of technologies and their relations.

AI can be understood as an 'umbrella term'[3] that includes a diverse set of technologies (see Figure 1 and for key terms, see Box 1), methodologies and applications aimed at enabling machines to carry out tasks typically associated with human intelligence such as learning, reasoning, problem-solving, and decision-making. In essence, there is a nested family of algorithms and computational approaches that fall under the category of 'Artificial Intelligence', from machine learning, via deep learning and neural

2 United Nations General Assembly. (2015). *Sendai framework for disaster risk reduction 2015–2030*.

3 European Parliament (2020). *Artificial intelligence: how does it work, why does it matter, and what we can do about it?* Luxembourg: European Parliament. Directorate General for Parliamentary Research Services. https://data.europa.eu/doi/10.2861/44572

networks, to generative AI. Importantly, AI has close links to expert systems and robotics. While certainly relevant (for example, in the context of drones), they are not the focal point of this Report.

Box 1. Key terms in AI for emergency and crisis management

- **Agent-Based and Multi-Agent Simulations**
  Computational models that simulate the actions and interactions of autonomous agents (individuals, households, organisations) to assess system-level outcomes (e.g. evacuation dynamics).
- **Agentic AI**
  AI systems with autonomous decision-making capabilities (Acharya, Kuppan & Divya, 2025).
- **Artificial Intelligence (AI)**
  An umbrella term for computational methods that enable machines to perform tasks typically requiring human intelligence such as learning, reasoning, perception, and decision-making.
- **Computer Vision (CV)**
  AI methods that enable machines to interpret and analyse visual data (e.g. photos, videos, satellite images). In crisis management, CV is used for tasks like flood mapping or damage detection.
- **Deep Learning (DL)**
  A branch of machine learning (ML) that uses multi-layered artificial neural networks to model complex, high-dimensional data patterns. It is especially effective for images, text, and speech.
- **Digital Twins**
  Virtual representations of social, physical and environmental systems, such as infrastructure, assets or even entire cities, that are continuously updated with real-time sensor data (Calderelli et al., 2023). The European flagship technology, *Destination Earth*, is promising here (Hoffmann et al., 2023).
- **Foundation Models**
  Very large, pre-trained AI models (e.g. in text, images, multimodal data) that can be adapted ('fine-tuned') to many downstream tasks, including emergency contexts.
- **Generative Models**
  AI models that can create new content based on training data, e.g. producing text, images, code or simulations. Applications include generating scenario narratives or structured protocols.
- **Large Language Models (LLMs)**
  AI models trained on massive text corpora, capable of 'understanding' and generating text-based content. LLMs are used for summarising reports, chatbots, or extracting information from unstructured documents. LLMs fall under the broader category of generative AI.

- **Machine Learning (ML)**
  A subset of Artificial Intelligence (AI) that enables systems to improve performance on tasks through data-driven learning, rather than explicit programming. It includes supervised, unsupervised, and reinforcement learning.
- **Natural Language Processing (NLP)**
  The processing of natural language information by a computer. Related to information retrieval, knowledge representation, computational linguistics, and more broadly with linguistics (Eisenstein 2019).
- **Rule-Based Systems/Expert Systems**
  A specific type of expert system that applies explicitly programmed with 'if–then' rules to reach conclusions or trigger actions. While important in emergency protocols, such systems are not considered AI unless combined with learning or reasoning capabilities.
- **XAI or Explainable AI**
  AI techniques that have the explicit goal of allowing humans to understand the underlying explanatory factors of why an AI or ML suggestion or decision has been made (Dwivedi et al., 2023).

The essential feature of AI in emergency and crisis management is to improve situational awareness: the capacity to perceive, plan for, and assess evolving environments accurately, which AI supports through data fusion and dynamic situational modelling (Comes, 2024; Endsley, 2017). Predictive and anticipatory capabilities are increasingly vital to decision support, enabling early warning and the forecasting of hazards like floods. The types of AI involved can include machine learning (e.g. predictive models for floods), natural language processing (e.g. chatbots for crisis communication), computer vision (e.g. aerial image analysis), multi-agent systems (e.g. evacuation simulation), and knowledge representation (e.g. structured reasoning over rules and protocols) (Comes, 2024 Kox, Harrison, Ziegler, & Gerhold, 2025; Lee, Comes, Finn, & Mostafavi, 2022).

AI has become a valuable means to collect and analyse heterogeneous and rapidly changing data (Nunavath & Goodwin, 2019; Kuglitsch, Pelivan, Ceola, Menon, & Xoplaki, 2022), promising several advantages over traditional models or human reasoning. These include the speed of data processing, improved temporal and spatial accuracy and detection of complex patterns (Reichstein et al., 2025).

Crisis data is extremely heterogeneous. If infrastructure is destroyed, or access is poor, there remain problems of sparse situational data, for example, regarding people in need or access conditions. At the same time, there are vast amount of data that can now be produced and accessed remotely, ranging from communications via social media and crowdsourcing, to satellites and drones that provide imagery. All these data are generated at high speed, and with varying degrees of veracity.

However, these data streams may only provide *proxies* for the information needed for preparation or response. For instance, satellite images may show the extent of flooding in an area, but not how many people are trapped without access to safe drinking water. UAV (drone) footage can reveal the size and spread of a wildfire, but not how much fine particulate matter is in the air that may pose a health risk to nearby communities.

Box 2 provides key definitions:

Box 2. Definitions for emergency and crisis

Definitions for 'emergency' and 'crisis' appeared in the *SAPEA Evidence Review Report* (SAPEA, 2022). The terms are used flexibly throughout this *Rapid Evidence Review Report*.

- **Emergency**
  An emergency is an imminent, serious situation requiring immediate action. It tends to occur with some sort of regularity, allowing professionals to prepare a response to particular types of emergencies.
- **Crisis**
  A crisis occurs when people perceive a severe threat to the fundamental values or functioning of a society or system, requiring an immediate response that must be delivered under conditions of (deep) uncertainty (Boin, Ekengren, & Rhinard, 2016; Rosenthal, Charles, & Hart, 1989, as cited in SAPEA, 2022).

Defining the use of AI in crisis management

The use of AI for crisis management can be defined by certain boundaries that help delimit its scope and provide focus. These boundaries are defined by: (1) its **function and context** (including operational boundaries) (2) its **methods** (the use of AI-specific technologies) and (3) its **limits** (including ethical, governance, policy and framework boundaries).

AI for emergency and crisis management can be framed across several dimensions, as outlined in Box 3. These can be influencing or determining factors that affect the role AI might play in such a situation and will be determinants in choosing which AI applications and tools are adequate to support the tasks at hand.

Box 3. Framing AI for emergency and crisis management

- **Temporal.** Real-time, anticipatory or retrospective support; prevention and preparedness, response or recovery phases of a crisis/emergency
- **Spatial.** Local (e.g. a building fire), regional (e.g. wildfire, floods) or global (e.g. pandemics)
- **Stakeholders.** Can include citizens, responders, decision-makers, non-governmental organisations (NGOs), other international organisations
- **Data ethics and governance.** Consent and privacy, data sharing prior to emergencies (for model training) or during emergencies, explainability and accountability
- **Socio-technical framing.** AI as part of a broader human-machine network, not as a replacement for human agency
- **General versus specific.** AI can be a generic model, addressing different types of questions (e.g. LLMs). or be designed and trained for a specific purpose
- **Users.** Can include operational responders/frontline workers, analysts, information management officers, domain experts, scientists
- **Maturity.** Established (e.g. satellite image analysis, weather forecasts) or experimental AI technologies and applications (e.g. agentic AI to support decision-making).

Functional and contextual boundaries are defined primarily by AI's purpose, that is, supporting humans in handling emergency/crisis situations. These are events that are time-sensitive, high-stakes, and uncertain. The use of AI must contribute directly to understanding, anticipating, mitigating, responding to or recovering from crises. Functional boundaries therefore include[4]:

- Early warning and forecasting (e.g. wildfire prediction)
- Real-time situational awareness (e.g. crowd monitoring)
- Impact and needs assessment (e.g. damage detection, including compound events)
- Information filtering and summarisation, preferably in natural language (e.g. social media analysis)
- Planning (e.g. evacuation routing)
- Resource optimisation (e.g. scheduling debris removal)
- Reporting (e.g. using LLMs for situation reports)
- Crisis communication (e.g. chatbots and agentic AI)
- Crisis training and preparedness (e.g. automated scenario design).

In defining **operational boundaries**, AI tools must be usable and robust in crisis environments, which are often resource-constrained and chaotic. Tools must be practical, interpretable, and resilient in real-world emergency settings, particularly with potential disruptions to information and communication infrastructures.

4 Functional boundaries may exclude generic AI applications that are unrelated to crisis/emergencies, even if used by emergency actors (such as human resource AI for hiring staff).

Methodological boundaries determine that it is not just any digital tool used during a crisis that can be understood as AI for crisis management. Rather, it must involve AI-specific methods and computational intelligence, such as learning, reasoning, or perception (not all computational methods are AI-based). The methodological boundaries therefore include, for example (see definitions in Box 1):

- Machine learning (supervised, unsupervised, reinforcement learning)
- Natural language processing (e.g. summarising situational reports)
- Computer vision (e.g. satellite image classification)
- Agent-based and multi-agent simulations (e.g. evacuation modelling)
- Generative AI (e.g. Large Language Models)
- Agentic AI (e.g. used for chatbots and automation).

The methodological boundaries exclude, for example:

- Static rule-based systems (see Box 1), without learning or adaptation
- Traditional GIS systems used purely for visualisation
- Manual dashboards and forms without AI-enhanced analytics.

While many applications also involve AI in robotics (for example, robots for search and rescue or unmanned aerial vehicles (UAVs) for remote sensing), these are not central to the scope of the present discussion. To acknowledge their role nevertheless, a box on the use of UAVs is included (see Box 4).

Ethical and governance boundaries (see systematic review by Batool, Zowghi & Bano, 2025) can be framed by a set of non-negotiable ethical or governance constraints that distinguish AI for emergency and crisis management from general AI systems. In short, the use of AI must uphold human dignity, transparency, and responsibility, even under pressure.

Policy and framework boundaries determine that an AI tool must align with international standards of safety, ethics, and crisis governance, such as the Sendai Framework (UNDRR), the EU AI Act, the OECD Principles on AI, and the IFRC/UN Guidelines. The degree of dependence on AI developed outside the EU is also an important aspect to consider.

AI tools for crisis management

Effective AI tools[5] in crisis contexts must be usable by a range of practitioners such as emergency responders, analysts, coordinators, and volunteers, who may lack deep technical expertise of such tools. As stated, AI tools also need to reflect the context of their use; while analytics for disaster preparedness may provide time for data collection, computation, and deliberation, their use during an operational response – especially in the field – requires information to be processed rapidly, and the technology needs to be rugged and robust. Operational applications include AI to support decision-making or

5 AI tools are applications or software that help users perform tasks and processes, communicate, or process data, or access information using a diverse set of AI technologies (see Box 1).

automate tasks, designed with simple interfaces, guided workflows, and with no technical expertise required. Such AI should reduce human workload and enable responders to focus on decision-making and frontline tasks (enhanced situational awareness; see Kox et al., 2025). For practitioners without deep technical expertise, AI tools must offer intuitive interfaces, easy integration, and clear outputs. Such tools may include[6]:

- **Crowdsourcing and data collection tools**, such as mobile apps or web forms for public or field data input (e.g. crisis mapping platforms), which allow non-experts to contribute to geospatial situational awareness through mobile or web-based crowdsourcing tools. These systems leverage participatory mapping, visualisation, and computation in a user-friendly manner (Middleton, Middleton & Modafferi, 2014). They emphasise accessibility and mobilise collective intelligence. Users do not need to understand the underlying ML algorithms; instead, they can focus on contributing and interpreting the outputs in practical terms (United Nations Office for Disaster Risk Reduction & CIMA, 2024).
- **Conversational tools and systems powered by large language models (LLM), or agentic AI,** such as chatbots or voice assistants for information access, retrieval or sharing. These may include AI platforms and decision support systems such as dashboards or risk maps that compile sensor, geospatial, and crowd-sourced data, providing AI-based recommendations, alerts, or digestible situational summaries to responders and real-time decision support (Acharya, Kuppan, & Divya, 2025; Betke, Peitzsch, Boldt, Reimann, & Kox, 2024; OCHA, 2021). Such models could enhance early warnings, assist with mitigation of misinformation, and support humanitarian coordination via natural-language summarisation and dynamic planning support (Odubola et al., 2025).
- **Training and simulation tools**, such as AI-based learning environments for crisis preparedness. Practitioners describe AI-enhanced simulations, especially for firefighter scenarios, as a powerful means for honing situational judgment under stress (van Leeuwen, Gasaway, Spaling & Netage, 2022). AI-powered simulation environments allow responders to practise high-stakes scenarios, such as de-escalation or crisis communication, in simulated but realistic contexts ('simulation-based training') (Pretolesi, Zechner, Guirao, Schrom-Feiertag, & Tscheligi, 2023). These systems enhance emotional preparedness, team coordination and decision-making under pressure, without exposing participants to real-world risks.
- **Automated analysis tools** that analyse data like images, text, sensor data, and/or reports (e.g. risk monitoring (Kikon & Deka, 2022), providing early warning (Zhou et al., 2024), damage detection (Yao et al., 2020; Voigt et al., 2007) and/or automate labour-intensive tasks. As an example, AI systems can support medical emergency call handling (Maletzki, Elsenbast & Reuter-Oppermann, 2024), or trigger alert notifications based on predefined criteria, thereby reducing responder workload and response lag. These tools also provide real-time dashboards combining sensor feeds, geospatial data, and social media, presenting clear warnings or action cues so that nonexpert users can make informed decisions swiftly.

6 Such tools do not include non-AI-based social media analysis (Reuter & Kaufhold, 2018), providing geographic information for disaster response through digital volunteers (Fathi & Fiedrich, 2022), drag-and-drop interfaces or no/low-code platforms, or purely visualisation platforms.

- **Hybrid architectures** that combine extractive and generative models offer adaptive, context-aware outputs - synthesising reports, summarising intelligence, and highlighting emerging hazards - via interfaces that non-experts can readily accept e.g. CrisisAI (Kazemi, 2025) and the ORCHID project (see section on Case Studies, below).

Uses and purpose of AI tools in crisis management

Requirements set out here draw on a broad range of reports on the use of AI in crisis management, such as the UNDRR report[7], the UNESCO readiness guidelines[8], and the OECD AI metrics catalogue[9], as well as the academic literature.

Generally, crises make information collection, analysis and decision-making difficult. Crises are characterised by time constraints, deep uncertainty, complexity, distributed authority, disrupted infrastructure, and demands for high levels of legitimacy and trust (Comfort, 2007; Mendonça, Beroggi, Van Gent, & Wallace, 2006; Muhren & Van de Walle, 2010; Paulus, Fathi, Fiedrich, de Walle, & Comes, 2024; 't Hart, Rosenthal, & Kouzmin, 1993). These conditions have also been shown to change how humans process information, make sense of their environment and make decisions – leading to a range of biases (Kahneman & Lovallo, 1993; Klein, Calderwood, & Clinton-Cirocco, 2010; Weick & Weick, 1995). Crisis conditions and their impact on human cognition, sensemaking and decision-making also shape whether and how AI can be used to improve information collection, analysis and decision-making across the different phases of crisis management.[10] Operationally, we lean here on the taxonomy for Crisis Information Management Systems (Tax-CIM) (see, for example, Borges et al. (2023) to outline requirements of AI tools in emergency and crisis management).

Tax-CIM defines seven dimensions mostly related to the response phase, based on previous work from its authors (Canos, Alonso & Jaen, 2004):

1. **Coordination**, including fieldwork coordination, command-to-fieldwork coordination, command coordination, coordination of the public, volunteer organisations, volunteers, resource management, logistics management, and adaptation on the fly
2. **Collaboration**, including role management, group decision-making, shared workspaces, and implementation of collaborative processes
3. **Information management**, including information capture and curation, maps, wearables, awareness of process and context, open data, data integration, information retrieval, public awareness, and decision logging

7 https://www.undrr.org/publication/documents-and-publications/special-report-use-technology-disaster-risk-reduction

8 https://www.unesco.org/en/articles/readiness-assessment-methodology-tool-recommendation-ethics-artificial-intelligence

9 https://oecd.ai/en/catalogue/metrics

10 The Report focuses on the preparedness and response phases, as requested by DG ECHO.

4. **Visualisation**, including customisation, desktop, mobile clients, tabletops, augmented reality, dashboards, and mashups
5. **Communication**, including videoconferencing, exclusive communication channels for responders and for control rooms, social media interaction with the public, one-way/two-way non-social-media interaction with the public
6. **Intelligence**, including recommendation and automatic decision-making, automatic information capture, filtering and categorisation, automatic inference[11]
7. **General support**, including robustness, privacy preservation, provenance, trust, scalability, interoperability, and open data.

A structured taxonomy such as this can help anchor diverse AI tools and functions within emergency and crisis management, guiding practitioners and researchers in understanding key functionalities and gaps for using AI. The scheme provides both a functional and a normative framework through which AI applications can be systematically assessed, implemented, and regulated in crisis contexts.

Preparedness phase

Using the Tax-CIM taxonomy, the following uses and purposes of AI are identified:

1. **Coordination and collaboration**, including prioritisation of vulnerable areas (Eini, Kaboli, Rashidian, & Hedayat, 2020), populations (Sirenko, Comes, & Verbraeck, 2025) or infrastructures (Esparza, Li, Ma, & Mostafavi, 2025), resource optimisation (e.g. AI-aided planning for shelters, stockpiles, logistics) or pre-positioning (De Clercq et al., 2025), and AI-driven training, exercises and simulation for responders (Conges, Evain, Benaben, Chabiron, & Rebiere, 2020; Pretolesi et al., 2023), including the generation of synthetic (images and other sensor) data for training or awareness that closely resemble real data recorded by sensors.

Such models with synthetic data are already used to create additional training data for AI. Synthetically generated data can also be used to perform interventions in the sense of Pearl's ladder of causality[12]. In other words, it enables machine learning models to generate output in response to input changes, without requiring the input to be recorded from real sensors. This is especially relevant as real interventions are, in many cases, either ethically or practically impossible. At the same time, it is essential to avoid overusing these models, since generative AI models have been shown to collapse over time if trained with recursively generated data (Shumailov et al, 2024).

2. **Information management**, including a) data collection (Middleton et al., 2014; Odubola et al., 2025; United Nations Office for Disaster Risk Reduction & CIMA, 2024), the integration of diverse sources (sensors, IoT, climate models, health records, historical disaster databases), data quality assurance (validation, noisy data), and b) data analysis and modelling (Asnaning & Putra, 2018; Pota

11 Please note that Tax-CIM Dimension Intelligence refers to the entire topic of automation.

12 See https://web.cs.ucla.edu/~kaoru/3-layer-causal-hierarchy.pdf (*What if? What if I do X? Example: What if I take aspirin, will my headache be cured? What if we ban cigarettes?*)

et al., 2022; Bhatia, Ahanger & Manocha, 2022), including hazard early warning and alert models (e.g. floods, landslides) and scenario simulation for resource allocation and evacuation planning.

3. **Visualisation and communication**, including customisable tools for visualisation (risk maps, infographics, dashboards), reporting, which address emergency managers, policymakers, communities, and other non-technical users (Middleton et al., 2014), and multi-lingual communication, i.e. the ability to translate alerts and other information material.
4. **General support**, including interoperability and the adherence to open standards (e.g. Common Alerting Protocol (CAP) for alerts), and data quality standards (such as accuracy, completeness, timeliness, consistency, relevance (Fisher & Kingma, 2001).

Response phase

Using the Tax-CIM taxonomy, the following uses and purposes of AI are identified:

1. **Coordination and collaboration**, including data and information sharing (Qadir et al., 2016), cross-agency exchange that offers secure, privacy-preserving and rapid data-sharing between emergency services, NGOs, and governments, and resource prioritisation to optimise resource allocation for rescue, and/or the location of relief hubs (distribution centres, field hospitals); relief distribution in real time and AI-assisted evacuation guidance, e.g. path optimisation for effective evacuation planning (Takabatake, Asai, Kakuta, & Hasegawa, 2025).
2. **Information management**, including a) data collection (Betke et al., 2024; Hayes & Kelly, 2018; United Nations Office for Disaster Risk Reduction & CIMA, 2024; Urbanelli, Frisiello, Bruno, & Rossi, 2024) using real-time recording of social media, field reports, satellite imagery, UAV data, and sensor information; crowdsourcing interfaces using mobile/web apps for citizens to report incidents, and b) data analysis and modelling (Chaudhuri & Bose, 2020; Zhang, Liu, Jiang, Fan, & Song, 2016; Yao et al., 2017; Voigt et al, 2007), such as rapid/automated damage assessment via image recognition (satellite/drone) to identify disaster needs (Pan et al., 2025), trend detection using NLP for emerging needs in social media feeds.

An important property here is adaptation to the effect of non-stationary timeseries data[13], out-of-distribution detection[14] and uncertainty quantification of the system's response, due to the nature of the data being processed (extreme events are rare, they are not covered well in training data, and there can be a distribution shift of data regimes under climate change).

3. **Visualisation and communication** (Vassell, Apperson, Calyam, Gillis, & Ahmad, 2016), including live situation/ incident dashboards, automated multi-channel dissemination via SMS, radio, social media bots, local language translations, and AI-generated summaries/briefs for responders and decision-makers; chatbots for public communication.

13 Data that changes due to trends or seasonality (see below for more information).

14 Data that differs significantly from the distribution on which a machine learning model was trained (see below also).

4. **General support**, including low-bandwidth adaptability using compressed formats for disaster areas with weak connectivity.

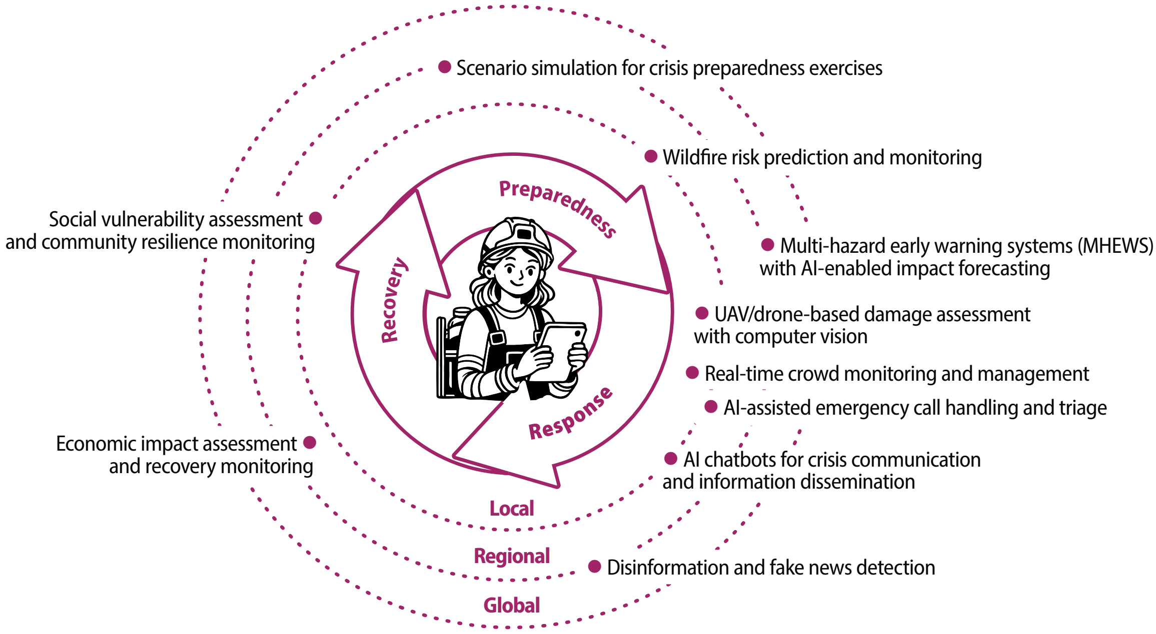


Figure 2. A taxonomy of AI Tools for crisis preparedness, response and recovery.

Figure 2 provides an overview on how the boundaries and considerations lead to a taxonomy of AI tools, both across the phases of crisis management and the spatial scale (from local to global). While the phase determines the time horizon for tool development, information collection, training, processing and decision-making, the spatial scale determines the need for contextualisation (locally), or the availability of interoperable high-quality datasets across regions and countries (at global scale). Figure 2 also showcases the diversity of AI tools and how they fall into different categories. While, for instance, AI-generated scenario simulations for crisis preparedness (especially in the context of the Union Civil Protection Mechanism (UCPM)) will happen at regional scale, following standardised protocols and with time to plan carefully, an operation like real-time crowd crisis management and monitoring typically happens very locally at dedicated event locations and requires contextualisation to determine and understand what constitutes abnormal behavioural patterns, preceding potentially dangerous situations.

3. The performance of AI in crisis preparedness and response

Introduction

Understanding task allocation and control

Crisis management authorities across the globe increasingly explore the use of AI for tasks such as monitoring and anticipation; assessing and reporting damages; and supporting or even automating decision-making (for example, under the umbrella of anticipatory action (Kjærum & Madsen, 2025). *Stanford's 2025 AI Index* report also highlights[15] that the technical performance of AI across several benchmarks continues to improve, at times significantly, even though complex (multi-chain) problems remain problematic. At the same time, there remains a certain scepticism about the use of AI, especially in sensitive and high-stake contexts such as disasters and crises (Sandvik, Jacobsen, & McDonald, 2017; Crawford & Finn, 2015, Bhatnagar et al, 2025).

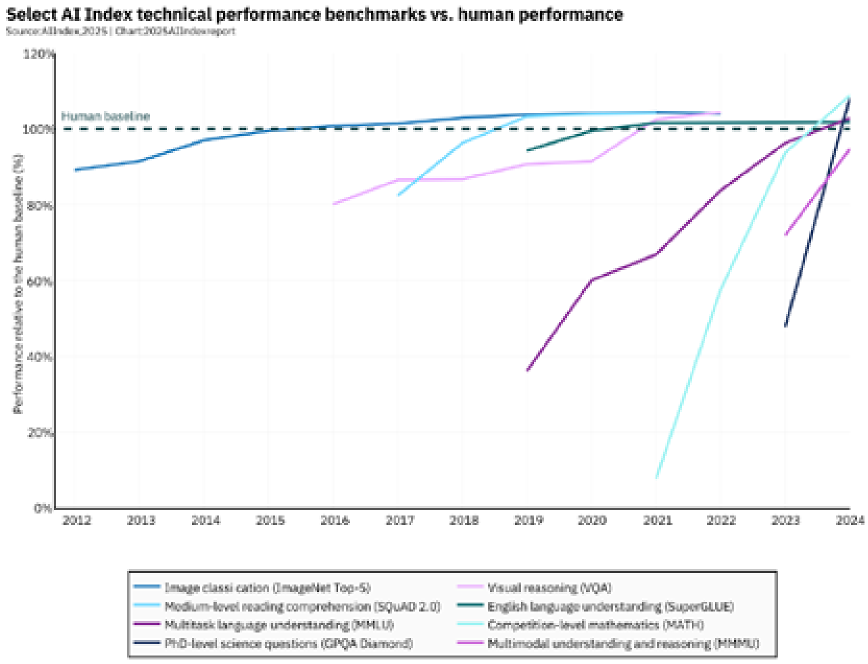


Figure 3. AI performance against selected benchmarks over time (Stanford AI Index report 2025).

The question of whether and how to deploy AI in crisis management requires consideration of two interrelated dimensions. Firstly, what tasks AI and humans each perform well, and secondly, how much control humans retain over AI-driven processes. While these dimensions of task and control are often

15 https://hai.stanford.edu/ai-index/2025-ai-index-report

discussed separately, it is useful to specify the required or desired control level, dependent on the different tasks at hand.

Whether AI helps depends on whether its capabilities fit the tasks at hand, and whether it is supporting existing organisational processes and protocols. In acknowledging the plethora of AI tools, applications, crisis decisions and tasks, combined with the rapid advancement of AI technology and research, this section of the Report does not attempt to assess the performance of each tool for every task. Rather, it provides guidelines that can help define requirements and assess whether and how AI can be useful. We first cover general considerations about 'outsourcing' tasks or functions to AI and automating them. We address the complexities of the following crisis management phases and functions: (1) monitoring, predicting, anticipating (2) assessing, reporting (3) decision support. We then briefly discuss guardrails and risk mitigation measures for the use of AI. Requirements are drawn from a broad range of reports on the use of AI in crisis management, such as the UNDRR report[16].

As a next step, we analyse how AI capabilities compare to human processes. Using the expanded HABA-MABA-AABA framework *(Humans Are Better At - Machines Are Better At - AI Is Better At)* (Bradshaw, Dignum, Jonker, & Sierhuis, 2012) developed from foundational work by Fitts (1951) and updated by Cummings (2014), we identify guidelines and guardrails for how humans and AI can collaborate in crisis management. Importantly, classic 'levels of automation' or task-transfer models (Endsley, 2017; Parasuraman & Riley, 1997) map cleanly onto the crisis information cycle of information acquisition, analysis, choice, and execution. However, they assume a single operator and stable context. In crises, many humans and machines interact; higher autonomy (i.e. outsourcing more tasks to AI) can reduce performance by degrading situational awareness and increasing coordination breakdowns, when interdependence is not designed in (Endsley, 2017). This is precisely why the HABA/MABA framework has evolved towards *hybrid intelligence* (Akata et al., 2020): designing for co-activity, common ground, observability and directability across human–machine teams, not simply reallocating functions to a human or to AI (or a robot). The ORCHID project (see Case Studies) provides an early example of such a framework and highlights the dynamic nature that is required in such hybrid systems.

Performance of AI in areas of crisis management

In the following section, we distinguish the three crisis management areas: monitoring and predicting, assessing, and decision-making. We discuss where AI performs well, where and why human intervention is crucial and how hybrid human-AI teams may work.

Monitoring, predicting, anticipating

For a range of hazards, AI can rapidly process large volumes of data and detect patterns. For instance, AI weather models' forecasts of the global weather, including aspects of extreme events such as tropical

16 https://www.undrr.org/publication/documents-and-publications/special-report-use-technology-disaster-risk-reduction

cyclones tracks, have been shown to outperform predictions from the best numerical models by up to 10 days (Sun et al., 2025; see also case study below). Machine learning approaches can process satellite imagery, data from sensor networks, and weather data at scales and frequencies that would be impossible for humans to achieve. In many cases, these ML-based models display a higher accuracy than conventional atmospheric models, as they can exploit statistical correlations beyond process-based theory (Reichstein et al., 2025). AI systems also show advantages in pattern recognition and predictive analysis when processing large historical datasets for hazard prediction (Jones et al., 2023). Areas of application include early warning systems (for instance, for floods (Zhou et al., 2024) or droughts (Kikon & Deka, 2022)), forecast-based financing (Coughlan de Perez et al., 2015), and food-security prediction (Deléglise et al., 2022). The performance of AI models is best for repetitive hazards that occur frequently, in similar or comparable contexts. Traditionally, AI has not performed well when confronted with new situations that fall outside the scope and context of its training, potentially missing unprecedented threats or low-probability, high-impact events. However, there are advances with respect to predicting rare *grey swan events*[17] for tropical cyclones (Sun et al., 2025), which may also apply to other extreme weather events.

In particular, **non-stationarity in data**, where patterns and relationships change over time, can be a challenge for using AI in crisis management. Non-stationarity of data distribution (such as timeseries, remote sensing images) generally arises from climate change, but also from the effects of other human actions on our Earth, both locally and globally. Responses to crises might also influence near-future data distribution. AI models trained on historical data may find it difficult to produce reliable or even sensible results on future data. Another important factor is that crises, by definition, are rare events, meaning that AI may have insufficient data from them. Consequently, crises can fall outside the known data distribution, risking AI responses that are incorrect. AI models need to incorporate as much domain or physical knowledge as possible, such as cause-and-effect relationships between observed variables, to help mitigate these problems. AI models can be highly valuable when integrated into ensemble models[18], together with other existing approaches. If an AI forecasting model does not significantly outperform alternative models, then combining them can enhance overall accuracy and robustness. This ensemble approach was popular in the past and is gaining more attention now, as different models often have complementary strengths and weaknesses. For instance, where AI models can sometimes be 'surprised' by unprecedented events that have not occurred in historical data, traditional forecasting models often excel at extrapolating forecasts for future events that differ from past patterns. Independent assessments, made with different models, are important to getting a more realistic estimation of uncertainty and to improve forecasts (see, for example, Wagenmakers et al, 2022).

In some situations, current state-of-the-art AI models may struggle to provide calibrated confidence in their results (see, for example, Guo, Pleiss, Sun, & Weinberger, 2017; Kendall & Gal, 2017; Venkataramanan, Bodesheim, & Denzler, 2025). They may tend to be overconfident, especially during erroneous decision-making or outputs caused by non-stationary data and potential distribution shifts. In crisis management, humans must always be aware of the stochasticity (inherent randomness) and

17 'Grey swan' events have some level of predictability, whereas 'black swan' events do not.

18 Ensemble models combine multiple individual models.

uncertainties in model outputs and the potential for AI errors. Methods and tools that enable AI to deliver calibrated levels of confidence are essential for deploying these models, particularly in human-in-the-loop scenarios. Such calibrated uncertainty forms the basis for recognising when AI is outside its familiar data regime. Such tools would enable users to trust AI outputs that struggle to return reliable or even reasonable results on future data.

Various authors also stress that context is critical in crises (Comes, 2024; Comfort, 2007; Mendonça, Jefferson, & Harrald, 2007; Sandvik, Jumbert, Karlsrud, & Kaufmann, 2014). While there are many calls around contextualising AI (Benaben et al., 2020), for now, humans themselves need to bring this contextualisation. This explicitly includes interpretation of anything AI flags as an anomaly, and monitoring whether the objectives and goals the AI is pursuing match the intention, thereby avoiding model drift. It is especially important when using LLMs with selective bias, which have been shown to lead to 'self-enforcing filter bubbles' (Sharma, Liao, & Xiao, 2024). Moreover, human capabilities remain essential for collecting contextual information that requires local knowledge, cultural understanding, and/or community engagement (Kox et al., 2025; Van de Walle & Comes, 2015). Crisis monitoring often requires information from informal sources, community observations, and tacit knowledge about local vulnerabilities that cannot be captured through sensor networks alone.

Assessing and reporting

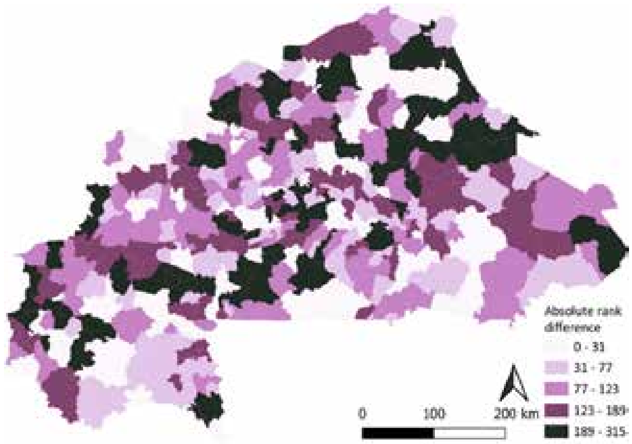
AI has advantages in rapid damage assessment and situation monitoring, particularly through the fast processing of large volumes of data, for example, via computer vision and/or spatial analysis of satellite imagery and UAV data (Gupta & Shah, 2021). Machine learning can process visual information that is especially difficult to manage, at speeds and scales that enable near real-time assessment of infrastructure damage (Xu, Lu, Cetiner, & Taciroglu, 2021) and/or population displacement (Tondaś, Kazmierski, & Kapłon, 2023). AI has the potential to enable multi-hazard risk assessment, addressing the interrelationships between hazards (Zhang et al. 2023), even for compound and cascading events. Social media analysis by AI systems can be used to collect situational information from distributed sources (Palen & Anderson, 2016). However, access, polarisation and misinformation remain critical issues. AI systems are also good at data fusion, especially for standardised datasets that provide interoperability (Migliorini et al., 2019). Furthermore, large language models (LLMs) have been positioned as a way forward to complement citizen science and/or participatory projects for mapping, such as Open Street Map[19] and damage detection with multi-modal large language models (See et al., 2025). These LLM-based methods, however, still need to be tested further and validated.

AI is also increasingly used to assess resilience to different hazards or of different systems (e.g. Mandal et al., 2024; Xi & Mostafavi, 2025) or vulnerability (Zhang, Wang, & Lu, 2023; Yokoyama & Takefuji, 2026). Some authors argue that the unprecedented availability of data enables new insights into the patterns that drive resilience and vulnerability (Yabe, Rao, Ukkusuri, & Cutter, 2022), especially in the social domain (Mandal et al., 2024). However, given there is no objective ground truth (i.e. direct empirical evidence) in assessments of vulnerability or resilience, these models are harder to

19 https://www.hotosm.org/

compare with conventional assessments, which are often indicator- or case-study driven (Jones et al., 2023). Furthermore, there are no standardised datasets, (machine learning) approaches or even data processing standards that facilitate comparison across contexts, for instance, in urban spatial data analyses (Casali, Aydin, & Comes, 2022).

While there is often discussion around the underlying datasets, methods also matter. While indicator-driven methods like INFORM[20] use theories to establish relations between different variables and social vulnerability, AI-driven models aim to derive these relationships directly from the data. For social vulnerability, there is comparative research that points to important differences between indicator- and ML-based analysis for dynamic contexts, such as in Burkina Faso, where an analysis of INFORM versus a PCA-based social vulnerability assessment reveals stark differences, even though the same datasets are used (Savelberg, Casali, van den Homberg, Zatarain Salazar, & Comes, 2025). Here, important considerations also need to be made in terms of explainability, theoretical grounding, contextual interpretation and co-creation; see Figure 4 below.



Category	Aspect	Inductive: SoVI	Hierarchical: INFORM
Selection of indicators	Choices	Automated	Theory-driven
	Contextualisation	Automated	No, global standard
	Accounting for double counting	Yes (+)	No (–)
	Large numbers of data sets	Possible	Max 10–20
	Data requirements	Very high	High
Dynamic behavior represented	Spatial consistency	Yes (+)	Yes (+)
	Temporal assessment	Not possible for Burkina Faso (–)	Possible (+)
	Interpretability	Complex (–)	Based on literature (+)
Suitable for decision making	Computing time	Time consuming (–)	Quick (+)
	Black box	Yes (–)	No (+)
	Humanitarian principles	No (–)	No (–)
	Intrinsic functioning	Medium (–)	Good (+)
	Post-hoc evaluation	Good (+)	Good (+)

Figure 4. Absolute rank differences between indicator and ML-based social vulnerability assessments of the communes of Burkina Faso highlighting that many communes that are the most vulnerable with INFORM are the least vulnerable based on PCA, and vice versa. *Table: Comparison of approaches and implications* (Savelberg et al., 2025).

20 https://drmkc.jrc.ec.europa.eu/inform-index

Information collection for crisis assessment therefore requires critical evaluation of source reliability, data quality, and potential bias, which humans perform more effectively than current AI systems (Crawford & Finn, 2015). The literature suggests there are limitations in AI's capacity for contextual interpretation and quality assessment of complex information. This is problematic, as key crisis documents such as humanitarian situation reports have been described as 'fundamentally confused' (Finn & Oreglia, 2016) and are thus far from standardised. The verification and validation of information sources, which is particularly critical in conflict situations or when dealing with potentially manipulated information, requires human judgement that can assess credibility, detect inconsistencies, and identify disinformation campaigns. Human analysts may be better at identifying when information requires additional verification, or when situational changes invalidate earlier assessments, or when local knowledge contradicts the suggestions of AI, especially in complex contexts such as conflicts (Van de Walle & Comes, 2015).

Decision support

AI has clear advantages in collecting and processing vast amounts of information and recognising patterns (Duan, Edwards, & Dwivedi, 2019). On the basis of such data, optimisation tools or scenario analyses, which are increasingly popular in combination with machine learning (Grass, Ortmann, Balcik, & Rei, 2023), can be used for various decision-support problems and can factor in trade-offs between multiple objectives and/or complex constraints, for instance, for location-allocation problems (Tanti, Efendi, Lydia, & Mawengkang, 2022). Importantly, the step from assessment and situational awareness to allowing AI to take decisions requires reflection on decision authority and control (see below).

Decision-relevant information collection often requires tacit knowledge, institutional understanding, and relationship awareness that cannot be captured in AI (Comes, Van de Walle, & Van Wassenhove, 2020). The political and institutional dimensions of crisis decisions in and across Europe require contextual information about policies, volatile power dynamics and institutional relationships that AI may not be able to provide.

For analysis and decision-making, LLMs are used increasingly for rapid feedback or as a 'co-pilot' on crisis decisions, for instance in crisis communication, or for resource optimisation (Odubola et al., 2025). The literature on crisis applications of LLMs is still sparse, so we draw here on general findings on the use of LLMs. Today, LLMs are still prone to 'hallucinations', i.e. generating output that sounds plausible, but is actually wrong and misleading (Huang et al., 2025). Some even expect these hallucinations to be an ongoing feature of LLMs (Banerjee, Agarwal, & Singla, 2025), and therefore will require permanent fact-checking and oversight, possibly supported by better explainability (Gunning et al., 2019). The reinforcement learning mechanisms that many LLMs use to incorporate human feedback have been found to provide limited diversity of output, and over-alignment to human preferences (Chaudhari et al., 2025). Both are problematic in crises, because of the importance of extreme or outlier scenarios.

Furthermore, an over-reliance on LLMs can lead to an erosion of critical thinking and, over time, of the skills and situational awareness of the analyst or decision-maker using the LLM (Crowston & Bolici, 2025). Particularly in crisis decisions, where moral trade-offs may be prominent, the potential of *moral deskilling* should also be considered if hard choices are routinely delegated to or supported by LLMs (Vallor, 2015). LLM echo chambers (Sharma et al., 2024), in combination with confirmation bias (Paulus et al., 2022), can amplify the tendencies of decision-makers to overlook or discard potentially important information if it does not fit the current mental framework.

To communicate decisions or recommendations, chatbots are used increasingly (Piccolo, Roberts, Iosif, & Alani, 2018; Urbanelli et al., 2024), for instance, as a way to deal with the issue of multiple languages (Vanjani, Aiken, & Park, 2019). Such chatbots are designed to be emotionally intelligent (Bilquise, Ibrahim, & Shaalan, 2022), and also go from 'listening' within social media to a bi-directional exchange. However, there are persistent difficulties of polarisation and radicalisation (Bleick, Feldhus, Burchardt, & Möller, 2024), as well as the generation of echo chambers if the chatbots seek to create user attachment (Jacob, Kerrigan, & Bastos, 2025). Concerns about the gamification and 'technologising' of true human connection and care, for example, in the context of UAVs (see Box 4 below), also apply if human conversations are outsourced to chatbots.

Box 4. UAVs to the rescue: Drones in crisis management

Unmanned Aerial Vehicles (UAVs), or drones, were originally developed for military surveillance and reconnaissance. While their widespread use in military conflicts has gained prominence with the war in Ukraine, the use of UAVs has become more evident in crisis management (Wankmüller, Kunovjanek, & Mayrgündter, 2021). UAVs promise access to areas that are otherwise inaccessible due to destroyed infrastructure, or where there is potential danger to emergency services. They are also cheaper to source and deploy than traditional aircrafts or helicopters (Hoang et al., 2023).

UAVs today still serve two broad purposes[21]. Firstly, they can provide reconnaissance through imagery for monitoring dynamically evolving hazards such as wildfires, damage assessment, and/or victim detection. Second, they serve as vehicles to deliver much-needed cargo, such as medicines, to where they are needed. Although civil protection actors still predominantly use raw still or video footage for reconnaissance, UAV data have been shown to be well suited for computer vision-based processing that allows detailed 3D object modelling, and machine-learning-based identification of damage or victims (Nex, Duarte, Tonolo, & Kerle, 2019). UAV data can also support the recovery process (Ghaffarian & Kerle, 2019). Finally, UAVs can also be used to explore interior spaces of potentially hazardous structures, such as those damaged by earthquakes (Karam, Nex, Chidura, & Kerle, 2022).

AI plays an important role to steer, coordinate and process drone data, ranging from pattern recognition for imagery using neural networks (Islam, Rashid, Hossain, Fleming, & Sokolov, 2023) to audio-based search and rescue (Deleforge, Di Carlo, Strauss, Serizel, & Marcenaro, 2019). AI also allows for the coordination of multiple autonomous drones. Multi-agent control architectures allow swarms of UAVs to conduct distributed searches and adapt dynamically to changes in the environment. The military use

21 https://reliefweb.int/report/world/drones-humanitarian-action-guide-use-airborne-systems-humanitarian-crises

of UAVs has also highlighted vulnerabilities of drones, such as 'spoofing' (deceiving) GPS systems that may also be relevant for other types of crises.

At the same time, the use of drones demands reflection. In 2014, a report by UN OCHA looked at the use of UAVs in humanitarian response, noting that their use was particularly challenging in conflict settings, where it may be difficult for communities to distinguish drones potentially delivering assistance from those that pose a threat[22]. Secondly, as with the gamification of warfare, there are concerns about the "technologising of care" that may reduce human-to-human interaction and lead to a potential loss of dignity in crisis response (van Wynsberghe & Comes, 2020).

For the EU, the use of UAVs in crisis management therefore requires harmonised technical standards and legal frameworks for the deployment of drones in different types of crises; secure communication protocols to protect against adversarial attacks; and data sharing standards protocols that ensure interoperability, privacy and security.

Control frameworks are needed to understand the implications of using AI for decision-making and decision support. These frameworks distinguish different levels of human involvement in automated or AI-driven processes. We combine here the literature on autonomy (Nothwang, McCourt, Robinson, Burden, & Curtis, 2016), discussing which tasks and processes should be handed over to AI (or more broadly, a machine), and the design and ethics-oriented literature on meaningful human control that conceptualise AI as a social-technical system, discussing the principles that enable human control (Santoni de Sio & Van den Hoven, 2018).

Three primary configurations are widely recognised:

- In **human-in-the-loop** systems, humans are active participants. Decisions require human approval before AI can intervene. AI provides recommendations or analyses, but a human decision-maker must actively choose to implement them (Wu et al., 2022; Herrmann & Pfeiffer, 2023; Lettieri, Guarino, Zaccagnino, & Malandrino, 2023). This configuration seeks to maintain human decision authority. However, the literature on meaningful human control stresses that this is only feasible in so far as the AI recommendation or advice are not beyond the ability of the human to understand or question the advice (Calvacante et al., 2023).
- **Human-on-the-loop** systems allow AI to execute decisions within defined boundaries (the operating decision domain (ODD)). Humans act as supervisors who monitor AI and can intervene or override when necessary (Nahavandi, 2017). Examples here are often in the space of robotics and autonomous driving (Abraham et al., 2021). Human-on-the-loop systems are asking for human operators to detect when intervention is needed.
- **Human-out-of-the-loop** systems operate autonomously without routine human oversight, though humans typically retain ultimate authority to deactivate them. This configuration is, for instance, discussed in military applications (Trzun, 2024). The AI Act explicitly prohibits the use of

22 https://www.unocha.org/publications/report/world/unmanned-aerial-vehicles-humanitarian-response

human-out-of-the-loop systems for high-risk applications, including those relevant to emergency response (Article 14) (see also Box 5). Human oversight must be designed into the system and enable operators to monitor, understand, and intervene in the AI's functioning. Even in time-critical disaster contexts, this legal requirement rules out fully autonomous decision-making for high-risk tasks.

If AI is used to support decisions, then adequate levels of control, autonomy and oversight need to be defined for the AI systems to operate on. While there is a lot of work on establishing control, e.g. for weapons systems (Amoroso & Tamburrini, 2020; Ekelhof, 2019), automated vehicles (Calvert, Johnsen, & George, 2024) or in health (Hille, Hummel, & Braun, 2023), there appears to be no evidence as yet on adequate control levels for different applications in crisis management.

Where not to use AI

Identifying guardrails

The HABA/MABA framework identifies several areas where AI is inadequate for crisis management:

Morally challenging decisions and trade-offs should not be referred to an AI tool. Even though there are attempts to develop 'moral agents', there are fundamental criticisms of using AI for moral decisions (Van Wynsberghe & Robbins, 2019). While AI can help with tasks like sensing and information collection, information analysis, and risk analysis, the step from analysis to decision-making requires careful consideration. Many decisions in crises, ranging from the COVID-19 pandemic to large-scale droughts, heatwaves or wildfires, require decisions that deeply affect our values (European Group on Ethics, 2022) and involve fundamental trade-offs between competing interests and rights (Comes, 2024). Since many of the values are abstract and hard to formalise, AI cannot adequately represent them, nor can AI engage in democratic deliberation as a means to negotiate value conflicts. Moreover, it is questionable whether AI has a mandate to make or support such high-impact decisions. The literature on meaningful human control (Calvacante et al., 2023) suggests defining the moral operational design domain (moral ODD), so as to specify where and when a human-AI system can operate, along with a definition of the domain in which a system ought not or should not operate, from a moral perspective.

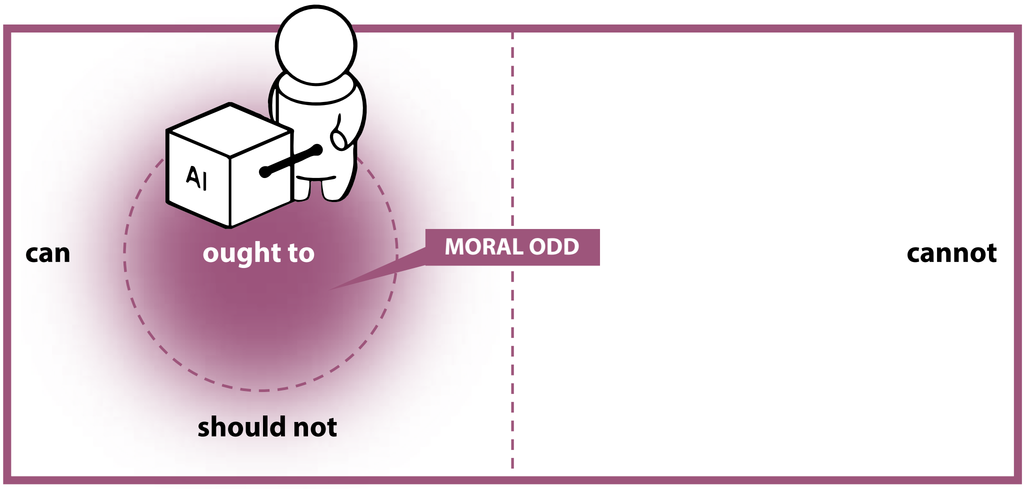


Figure 5. The principle of moral operational design domain: Human-AI system operates within the boundaries of what it can do and within the moral boundaries of what it ought to do (from Calvacante et al, 2023).

Context matters. Crisis management requires an understanding of local contexts, cultures, and community dynamics. These may vary significantly within Europe. AI systems trained on aggregated datasets or for one single context, hazard or region could struggle to capture another situation and may recommend interventions that are inappropriate or misleading. The problem of non-stationary data (or changing patterns) also requires a frequent recalibration of any AI applied to crisis management. Adjusting to context requires also that the distribution of roles and control authority between humans and AI ("who is doing what and who is in charge of what") is consistent with their individual and combined abilities (Calvacante et al., 2023).

New crises. Climate change or escalating conflict may bring about unprecedented situations, for which no training data are available. Yet, many extreme weather models assume explicitly or implicitly that future events will reflect historical risk. Machine learning currently excels at pattern recognition in historical data, even though new developments point to increasing abilities for transfer learning (recognising new contexts) in extreme weather events (Sun et al., 2025). The interpretive flexibility (Weick, 1995) required to *reframe* and rescope problems when initial approaches prove inadequate and which sensemaking theory identifies as essential for complex crisis management, remains beyond AI's capabilities.

Trust-building and relationship management require distinctly human capabilities (Lee & See, 2004). Building the interpersonal relationships and institutional trust required for effective European crisis coordination involves emotional intelligence, cultural sensitivity, and diplomatic skills that AI systems cannot replicate. The maintenance of solidarity and cooperation during challenging situations requires humans who can navigate competing national interests while preserving collaborative frameworks.

Figure 6 summarises the findings. Artificial intelligence can improve crisis preparedness and response. However, alternative approaches need to be used for moral decisions that require trade-offs; in situations where localised knowledge and contexts are important that cannot be transferred via data;

for unprecedented crises, for which there is no training data; and in situations where it is important to maintain and strengthen human connections and empathy.

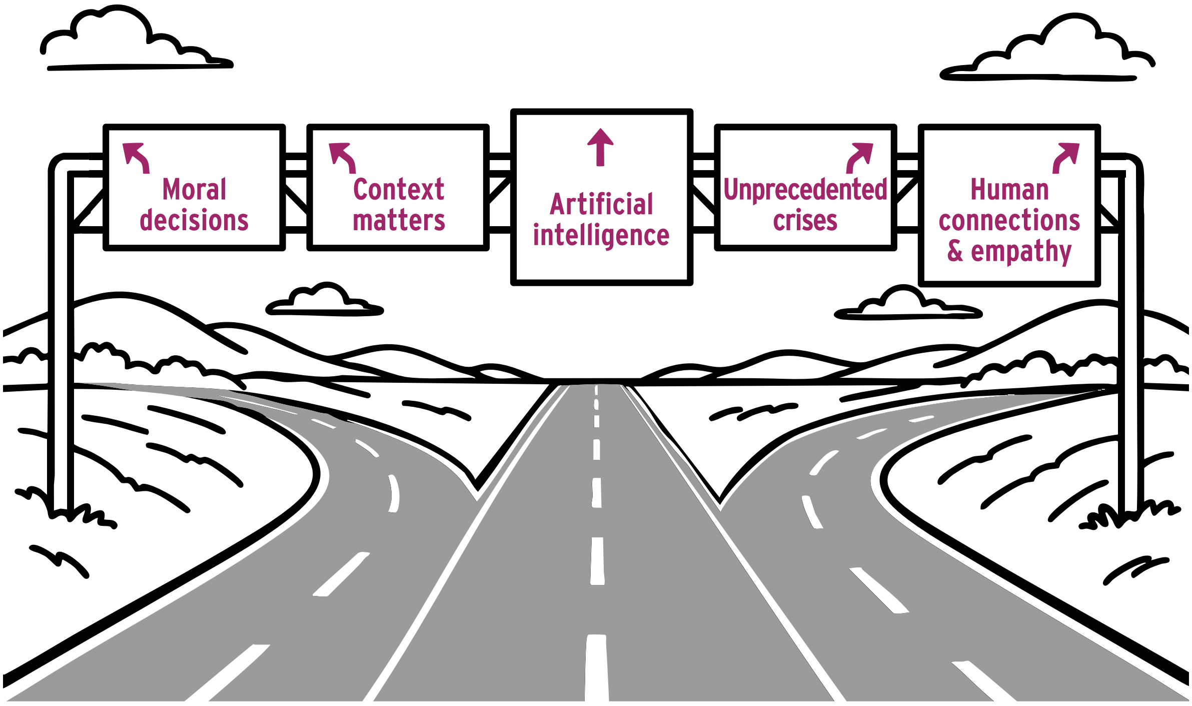


Figure 6. When to use AI for crisis management – and when to take alternative routes.

4. Legal and ethical requirements and guiding principles

Introduction

This section considers critical legislation, such as the EU AI Act and the GDPR, followed by 'soft law' ethical principles, frameworks and standards. It ends with two illustrative examples of the implementation of EU law in the context of crisis management.

The EU AI Act

The European Union's AI Act (Regulation (EU) 2024/1689) provides a comprehensive legal framework for AI. Both the AI Act and the GDPR have extraterritorial provisions that could be relevant for disaster and crisis scenarios, especially where AI systems or data processing cross borders (see Illustrative Examples below). The AI Act (pending full entry into application) applies to: (1) AI systems placed on the EU market or put into service in the EU, regardless of the provider's or deployer's location, and (2) providers and deployers of AI systems located outside the EU, if the output of the system is used in the EU. This means that non-EU organisations, such as US-based AI providers, must comply with the AI Act if their systems are used in EU disaster response, even if developed or hosted elsewhere. Similarly, EU-based AI organisations remain bound by the AI Act when deploying systems in the EU and globally.

The AI Act follows a risk-based approach, where AI systems are classified according to the level of risk they pose to health, safety and fundamental rights. Certain AI practices are prohibited, such as harmful manipulation, social scoring, and real-time remote biometric identification. High-risk systems need to comply with strict obligations before they are put on the market. These include systems intended to be used as, for example, safety components of critical infrastructures, or for the use of biometrics. They need to comply with obligations on, for example, robust risk management, high-quality data, transparency and human oversight.

European Commission guidelines that clarify the definition of an AI system and on prohibited AI practices were published in February 2025[23], with further guidelines to provide clarification on aspects such as the classification of high-risk AI systems expected in 2026. The classification of an activity as 'high-risk' (according to Article 6) means that the systems in question might be subject to strict obligations to ensure safety, transparency, human oversight and accountability, depending on entity type.

23 Commission Guidelines on AI system, https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application, and Commission Guidelines on prohibited artificial intelligence (AI) practices, as defined by the AI Act, https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-prohibited-artificial-intelligence-ai-practices-defined-ai-act

4. Legal and ethical requirements and guiding principles

The AI Act could potentially classify AI used in emergency and crisis response as 'high-risk' (according to Article 6, see also Box 5), meaning that such systems might be subject to strict requirements for safety, transparency, human oversight and accountability, depending on entity type. In cases where AI systems are classified as 'high risk' under the AI Act, such systems need to undergo conformity assessments and implement risk management, quality data training, logging, and human oversight mechanisms to ensure a high level of protection on health, safety and fundamental rights.

Box 5. Legal basis for classification of AI systems as "high-risk"

- Article 6(2): An AI system is classified as high-risk if it is intended to be used in a critical area listed in Annex III.
- Annex III 5(d): High-risk includes AI systems intended to "dispatch, or to establish priority in the dispatching of emergency first response services, including by firefighters and medical aid."
- Annex III does not currently include general-purpose disaster early warning, situational awareness, or risk prediction tools unless they directly affect the dispatching of emergency services or involve public benefits/services (Annex III (5)(d)).

Under Point 5(d) of Annex III of the AI Act, AI systems 'intended to evaluate and classify emergency calls by natural persons or to be used to dispatch, or to establish priority in the dispatching of, emergency first response services, including by police, firefighters and medical aid, as well as of emergency healthcare patient triage systems' are classified as 'high-risk'. It is therefore possible that certain AI systems used in crisis management will be classified as high-risk.

The classification of AI systems as 'high-risk' under the AI Act is not automatic, even for AI used in disaster or crisis management. It depends on the intended use of the system and whether the use falls under (a) one of the Annex III use case areas (Article 6(2) AI Act) or (b) it is to be used as a safety component of a product, or the AI system is itself a product covered by the Union harmonisation legislation listed in Annex I, which is required to undergo a third-party conformity assessment (Article 6(1) AI Act). Even for certain systems not classed as 'high risk', the AI Act provisions on General Purpose AI models and/or systems might still apply.

Use of general-purpose AI in crisis management contexts

While many AI systems used in crisis management may fall under the high-risk category defined in the AI Act (Article 6 and Annex III), an increasing number of tools deployed by emergency services (including language models, image analysis, or predictive tools) are based on General Purpose AI (GPAI) systems (e.g. a large language model (LLM) or foundation model). These are not developed for a specific use case but offer broad functionality across sectors. Under Title VIII of the AI Act, GPAI systems are subject to a distinct regulatory regime. If a GPAI model is used without being embedded in a high-risk application, it is not classified as high-risk itself. However, GPAI providers and deployers must still meet specific obligations, including transparency, technical documentation, disclosure of capabilities and limitations, and, for advanced systems, systemic risk mitigation.

Where GPAI is integrated into a high-risk application (e.g. for automated eligibility assessments or emergency dispatching), the provider of the final application bears the responsibility for fulfilling the high-risk system requirements (Title III), but GPAI developers must support with documentation and integration transparency (Art. 52(2)).

At the same time, the AI Act exempts, within the scope of Members States for military or national security purposes, AI systems used for these purposes (Article 2(3) AI Act). This means that if a crisis (especially CBRN incidents or terrorism) is handled by a Member State under national security or defence, those AI tools may fall outside the AI Act's governance − a gap that requires careful oversight to avoid loopholes (Gstrein, Haleem & Zwitter, 2024). For crisis management and disaster response, this means that civil protection operations remain within the scope of the AI Act, unless clearly operating under a national security mandate. Most civilian humanitarian and emergency uses of AI are therefore subject to the AI Act's provisions, including early warning, medical triage, resource allocation, or coordination support.

The AI Act in practice: Gaps, requirements, and operational implications

The AI Act establishes a risk-based legal framework for AI, placing the highest regulatory burden on 'high-risk' systems, many of which are relevant to emergency and crisis management. These include AI used in critical infrastructure, public services, life-critical decisions, and resource allocation, all common in disaster settings. While the AI Act defines broad risk categories, assessing whether existing tools meet these standards is still evolving and is highly context dependent.

To qualify as compliant under the AI Act, high-risk systems must meet the requirements in Title III, Chapter 2, including:

- Risk management system throughout the lifecycle
- High-quality, representative training and testing data
- Technical documentation and record-keeping
- Transparency and explainability of system function and purpose
- Human oversight mechanisms
- Accuracy, robustness, and cybersecurity
- Conformity assessment before market entry.

Yet, many AI tools currently piloted or used in crisis contexts, such as satellite-image-based damage assessment, LLM-based reporting assistants, or evacuation planning algorithms, do not yet fully meet these standards:

AI Act requirement	Observed gaps in crisis tools
Risk management system	Lacking continuous risk assessments and post-deployment monitoring
Data quality and representativeness	Training data often incomplete, biased, or not documented transparently
Transparency and explainability	Many models are 'black box' (e.g. deep learning, LLMs) with limited auditability
Human oversight	Human-in-the-loop is not consistently implemented or tested under stress
Technical documentation	Often informal or proprietary, lacking standardised documentation practices
Conformity assessment	Not yet in place for most tools used in humanitarian or civil protection settings

Table 1. AI Act requirements and observed gaps in crisis tools.

To align with the AI Act, particularly Articles 9–15, public authorities and developers must urgently adopt operational tools such as pre-deployment conformity assessments, ideally tailored to disaster use cases, and post-deployment audits, including stress-testing and red-teaming (adversarial testing)[24]. Developing a "Crisis AI Compliance Toolkit" at EU level could help standardise these practices.

General Data Protection Regulation (GDPR)

AI in crisis management often relies on 'big data' (for example, for geolocation of citizens, health or demographic data for evacuations or relief). The GDPR is a key legal pillar that ensures privacy and data protection in the EU. It requires that personal data used by AI is processed lawfully, for specific purposes, and with minimal necessity. There are emergency exceptions, and the GDPR does allow data processing in emergencies under certain bases. For example, processing may be lawful if it protects someone's "vital interests" or serves an important public interest in disaster situations. Recital 46 explicitly cites humanitarian purposes, such as monitoring epidemics or natural and man-made disasters, as a legitimate basis for data use in crises.[25] Even so, responders must ensure data minimisation, secure handling, and respect for individual rights. For instance, if AI analyses social media or phone data to locate survivors, it should use only necessary data and anonymise where possible, to comply with privacy principles. GDPR also enshrines transparency and the 'right to a human decision' in automated

24 'Red teaming' is a form of testing for potential vulnerabilities in systems.

25 Recital 46 - Vital interests of the data subject. *General Data Protection Regulation (GDPR)* [blog post]. Retrieved from https://gdpr-info.eu/recitals/no-46/

processing (Art.22 GDPR), meaning that individuals have the right not to be subject solely to automated decisions that significantly affect them. This is highly relevant in crisis aid scenarios; affected people should, where feasible, be informed about AI-driven decisions (like how aid is allocated) and have recourse to human review (McElhinney & Spencer, 2024).

Ethical guidelines and principles

Beyond hard law, a number of ethical frameworks guide AI usage in disaster management:

EU's Trustworthy AI Principles. The EU's High-Level Expert Group on AI has defined seven requirements for Trustworthy AI.[26] These include:

- **Human agency and oversight.** AI should empower human decision-making, not replace it, ensuring a human-in-the-loop for critical crisis decisions
- **Technical robustness and safety.** AI must be reliable and secure, with fallback plans to prevent malfunction in high-stakes situations
- **Privacy and data governance.** Strict protection of personal data and privacy throughout the AI system's lifecycle
- **Transparency.** The AI's logic, data sources, and outputs should be explainable to authorities and the public (which is important for maintaining trust in tools like risk predictions)
- **Diversity, non-discrimination and fairness.** AI should not worsen biases or exclude vulnerable groups, upholding fairness in, for example, resource distribution or evacuation planning
- **Societal and environmental well-being.** AI use should account for broader societal impacts and not harm the environment or societal cohesion
- **Accountability.** Clear responsibility must be established for AI outcomes, including auditability and the possibility of redress for harmful errors (Gevaert, Carman, Rosman, Georgiadou, & Soden, 2021).

These principles, grounded in fundamental rights, echo the core ethics of 'Do No Harm' and justice, and can be applied in different contexts.

Humanitarian principles and 'Do No Harm'. In disaster relief contexts, it is widely accepted that AI deployments must adhere to humanitarian ethics. This means prioritising human life, dignity, and impartiality. Any experimental use of AI ('*humanitarian experimentation*') on crisis-affected populations must be approached with caution and informed consent, where possible (Zwitter, 2018). There is growing concern that vulnerable communities should not become unwitting test subjects for unproven AI tools (Sandvik et al., 2017). 'Do No Harm' is paramount. AI should not put people at greater risk; for example, an algorithmic error should not mislead responders or deprive aid to those in need. Humanitarian guidelines call for *accountability to affected populations*; agencies should inform and consult communities about AI-assisted programmes, allow them to opt out of purely automated decisions, and incorporate their feedback. Transparency about how AI is used in aid (and its success

26 https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

or failure) helps maintain trust. Importantly, these ethical guardrails align with emerging legal norms; for instance, the right to a human review of AI decisions could become a standard part of EU law, protecting human agency in high-risk AI deployments.

In line with the AI Act's Article 5 (Prohibited Practices) and the ethics principles endorsed by the EU High-Level Expert Group on AI, the humanitarian principle of 'Do No Harm' should not be viewed as a regulatory constraint alone. Instead, it should serve as a core evaluative lens; any AI system intended for use in crises should be demonstrably *safe, fair,* and *context sensitive*.

'Do No Harm' criterion	Operational approach to demonstrate compliance
Avoid harm to individuals/groups	Conduct Algorithmic Impact Assessments focused on vulnerable populations
Fair and inclusive data practices	Ensure bias audits and demographic representativeness in datasets
Transparent decision logic	Use Explainable AI (XAI) and provide model explanations to users
Ethical alignment	Implement ethics-by-design protocols, using guidance from humanitarian data standards
Benefit-risk trade-off	Apply proportionality assessments: weigh predicted benefits (e.g. faster response) against potential harms (e.g. misclassification or exclusion)

Table 2. 'Do not harm' criteria and demonstration of compliance.

These methods go beyond compliance and support public trust, which is crucial in high-stakes emergencies. Where fully avoiding harm is impossible, tools should make trade-offs explicit, document uncertainties, and embed fail-safes and recourse mechanisms.

Global and sectoral codes of ethics. Existing codes of ethics, like the Core Humanitarian Standard (promoting accountability and community participation) and the UN's AI ethics guidelines, reinforce similar principles of fairness, accountability, transparency, and sustainability (Zwitter & Gstrein, 2020). The World Health Organization, in examining AI for public health emergencies, likewise concluded that strong ethical principles must guide AI use, to protect public trust and safety.[27] They warn that without careful governance, AI could introduce *algorithmic bias, privacy breaches, or exacerbate inequalities*. For example, if AI for risk communication targets the wrong groups or uses biased data, it may unintentionally harm vulnerable communities. The overarching theme is that ethical AI in crises

27 Responsible AI use can advance risk communication and infodemic management in emergencies, new study shows. (2025, May 23). *WHO News*. Retrieved from https://www.who.int/europe/news/item/23-05-2025-responsible-ai-use-can-advance-risk-communication-and-infodemic-management-in-emergencies--new-study-shows.

should be human-centric and rights-respecting, complementing human decision-makers rather than overriding them. As one EU expert noted, "AI is not a standalone solution. It must be embedded into operational workflows and linked to legal, ethical, and societal safeguards."[28] In practice, this means maintaining human oversight, conducting impact assessments, and ensuring that innovation never comes at the cost of human rights, trust or safety.[29]

Guiding frameworks and standards

To operationalise these legal and ethical requirements in crisis management contexts, several guidelines have emerged. The International Committee of the Red Cross (ICRC), with partners, published a *Handbook on data protection in humanitarian action*, (Kuner & Marelli, 2017) which translates data protection principles (like those in the GDPR) into crisis scenarios. It covers practical steps such as conducting Data Protection Impact Assessment (DPIAs) in humanitarian projects, obtaining consent in chaotic environments, and ensuring fair data processing for vulnerable subjects. Similarly, the UN OCHA's Centre for Humanitarian Data has issued detailed Data Responsibility Guidelines,[30] which emphasise data sharing agreements, role-based data access, and protection measures when humanitarian, government, and private sector actors collaborate on data during emergencies. These guidelines stress data minimisation, secure storage, and timely deletion once data is no longer needed, reflecting both the GDPR and the heightened duty of care owed to crisis-affected communities. They also introduce a data lifecycle approach to manage data from collection, through analysis to deletion, ensuring responsibility at each stage. Another notable framework is the *510 Data Responsibility Policy* from the Netherlands Red Cross' data science initiative [31], which enshrines principles like people-centred design, group privacy protection, and accountability for data use in disaster projects.

Illustrative examples: Legal frameworks in practice

To understand the practical implications of legal requirements during disaster scenarios, consider the following examples:

1. **Inside the EU – AI-supported flood response in Germany.** During the 2021 floods in Western Europe, AI tools were used to analyse satellite and sensor data for early warning and damage assessment. A French-based company provided predictive analytics services for regional authorities in Germany. As the AI system was deployed in the EU and involved the processing of geolocation and infrastructure data potentially linked to individuals, both the GDPR and the AI Act would apply today. The provider must ensure lawful data processing (e.g. under public interest or vital

28 *How can AI strengthen disaster preparedness in Europe?* (European Commission - UCP Knowledge Network), retrieved July 21, 2025 from https://civil-protection-knowledge-network.europa.eu/news/how-can-ai-strengthen-disaster-preparedness-europe.

29 See footnote 27.

30 *The Centre for Humanitarian Data. OCHA data responsibility guidelines.* Retrieved from https://centre.humdata.org/the-ocha-data-responsibility-guidelines/

31 https://510.global/wp-content/uploads/2025/01/Data-and-Digital-Responsibility-Policy-2024-V3.3.pdf

interest grounds), implement bias mitigation and human oversight, and comply with high-risk AI requirements under the AI Act.

2. **Outside the EU – EU-funded earthquake relief in Nepal.** In EU-funded humanitarian operations outside the EU, such as DG ECHO's support to earthquake-affected regions in Nepal, satellite-based AI is sometimes used to assess damage and guide logistical planning. If a European AI company provides these services, the AI Act still applies, since the system is 'put into service' by an EU actor. However, GDPR may not apply to personal data about non-EU individuals processed exclusively outside the EU, unless EU entities directly monitor individuals. Nonetheless, ethical data responsibility policies – such as those developed by the Red Cross or OCHA – are often used to fill these legal gaps, ensuring that fundamental rights are upheld, even where formal EU data protection does not apply.

Predefined decision-making rules and accountability under uncertainty

In crisis contexts, decisions are frequently made under conditions of deep uncertainty, where outcomes are probabilistic and evolving. AI can assist by accelerating data processing, offering predictions, or recommending courses of action. However, it must be emphasised that AI should not redefine the decision-making rules themselves. These rules, including thresholds for triggering action, principles of prioritisation, or loss functions used in resource allocation should be established ex ante, prior to crises, and remain independent of the specific technical tools employed (whether AI, expert judgment, or statistical models). This safeguards accountability, ensuring that decisions can be justified ethically, politically, and legally, even when outcomes are imperfect. The L'Aquila earthquake trial (Marzocchi, 2012) demonstrated the legal risks of failing to distinguish between uncertain forecasts and decision protocols. In this light, decision-making rules must incorporate both rational criteria (e.g. expected loss minimisation, precautionary principles) and societal values, such as fairness, proportionality, and the duty to protect. AI systems must be designed to work within these predefined frameworks, not to determine them, thereby reinforcing human oversight, legal defensibility, and trust in crisis governance.

Consideration of environmental impacts and energy use

Box 6 highlights a number of considerations of AI and particularly LLMs, in terms of environmental impacts and energy requirements.

Box 6. Environmental impacts of the use of AI, particularly LLMs

AI has the potential to save lives by accelerating response times and improving coordination across agencies. Yet, AI deployment comes with an environmental cost, mainly linked to energy consumption, use of water and raw materials, and emissions associated with training and running AI models (see Figure 7). The trade-off between mitigating human suffering versus adding to environmental pressures deserves explicit attention in policy discussions, especially as crises related to climate change become more frequent.

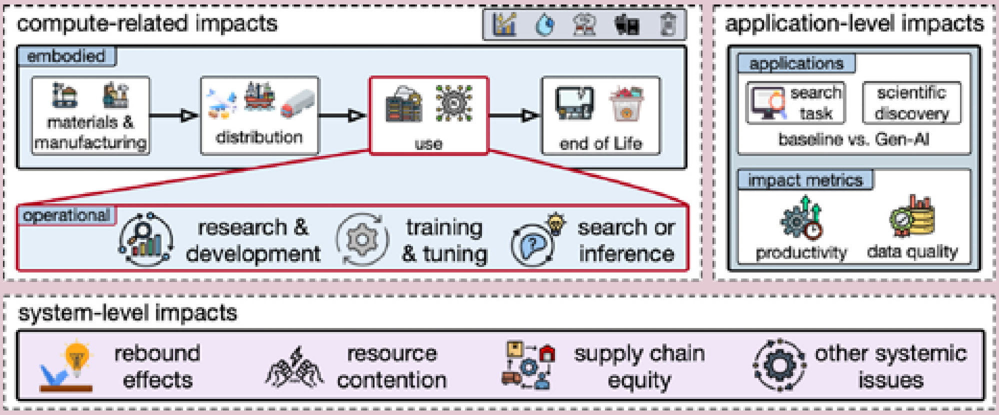


Figure 7. AI's environmental impact (from Bashir et al., 2024)

The carbon footprint of AI

Modern AI systems can consume huge amounts of electricity and thus contribute to greenhouse gas emissions, primarily due to the energy-intensive computations in model training and usage. The data centres that power the latest AI advancements are responsible for a significant amount of global electricity demand (1-1.5%, by some estimates[32]) and in some cases even exceeding the emissions of the aviation sector. AI's environmental impact can arise across different stages of its lifecycle: training, post-training (fine-tuning) and inference phases.

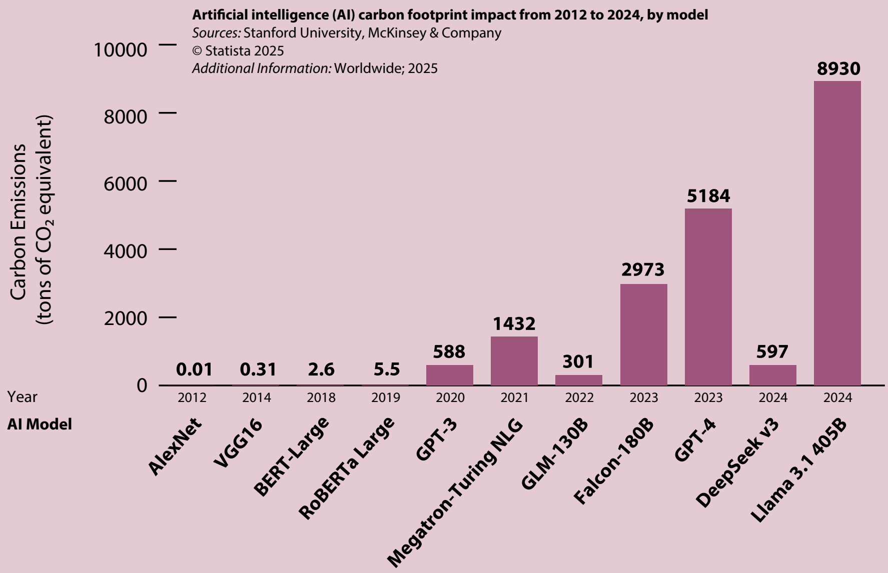


Figure 8. Carbon footprint of LLM training for different models.
Stanford University (2025), reproduced by Statista

32 https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

While training is performed a single time (or periodically), inference happens continuously, at-scale. A single short GPT-4o query consumes 0.42 Wh; scaling this to 700 million queries/day results in substantial annual environmental impacts (Jegham et al., 2025).

Strategies for efficiency

The principle for green algorithms is simple; more efficient code and algorithmic optimisation mean less computation is needed for the same task, which in turn means less electricity used (and less CO_2 emitted if that electricity is not 100% green). There are some key strategies and developments to achieve greener algorithms: algorithmic optimisation, efficient coding or using specialised hardware.

Implications for crisis management and climate action

Green AI is particularly relevant for crisis management, where AI systems are increasingly used to predict and respond to climate-related disasters. While these tools can save lives, they also consume large amounts of energy and water through data centres and power-hungry infrastructure, potentially worsening the very crises they aim to address. Currently, there is no data or evidence on the specific environmental impact of AI in crisis management, making it hard to gauge the sector's share of the overall footprint. It is therefore recommended that energy consumption and other resource metrics are logged during deployment and operation. Prioritising low-resource approaches, such as lightweight models running on renewable-powered or battery-operated devices, can reduce emissions and ensure operability in disaster zones with limited energy or connectivity. Aligning AI innovation with sustainability is essential to preventing new vulnerabilities and ensure that AI becomes part of the solution rather than an additional burden.

5. Data governance and sharing

Introduction

AI is only as good as the data it learns from and operates on. In European civil protection, relevant data (such as satellite imagery, weather data, social media feeds, sensor readings) comes from a myriad of sources, for example, EU agencies, national governments, private satellite companies, social networks, NGOs, and citizens.

Challenges of data governance and sharing

This section identifies several critical challenges that specifically pertain to crisis management in the EU, where cross-border crisis response requires collaboration and thus data sharing across multiple Member States.

Data availability and quality. Data scarcity remains a problem for certain hazards or regions, which is potentially amplified by data silos between institutions. AI models need large, representative datasets to be reliable, yet historical disaster data might be patchy or biased (for example, underreporting impacts on marginalised communities). Governance mechanisms like the EU's data spaces, the newly introduced Data Labs or emergency data hubs can facilitate the sharing of relevant datasets among authorised parties.

In addition to availability, data quality is essential. There are several frameworks that discuss data quality, most prominently the Generic Data Quality Assurance framework by the UN[33]. Information quality can depend on context, user and information perspective (Van de Walle et al., 2015). For crises in particular, a variety of frameworks have been proposed (Fisher & Kingma; 2001; Seppänen & Virrantaus, 2015; Van de Walle & Comes, 2015), which include attributes such as:

- **Accuracy.** Data conform to the real-world fact or value
- **Timeliness.** Data are not out-of-date and in time for a decision to be made
- **Completeness.** Data represent the phenomenon fully, and there are no structural biases and/or omissions
- **Consistency.** Data are free of contradictions
- **Relevance** refers to the applicability of data in a particular context
- **Accessibility** of data is continuous and guaranteed, and
- **Traceability** of data provenance and verifiability.

33 https://unstats.un.org/unsd/unsystem/Documents-Sept2015/GSQAF-GenericData-Sept2015.pdf

It is also possible to distinguish context dependent and independent information quality attributes. Intrinsic attributes focus on the data itself; problem-centred attributes on suitability to the task at hand; representational attributes consider ease of interpretation and understanding. Figure 9 below summarises the most important attributes for crisis management (Van de Walle & Comes, 2015).

Category	Context Dependent Attributes	Context Independent Attributes
Intrinsic	Credibility, Reputation	Accuracy, Objectivity
Problem-Centered	Value Added, Timeliness, Relevancy, Appropriateness	Completeness
Representation	Interpretability, Ease of Understanding	Consistent and concise representation

Figure 9. Information quality attributes (from Van de Walle & Comes, 2015).

To reflect these considerations of data quality, repositories like the humanitarian data exchange website have put forward their own data guidelines and quality trackers[34]. For AI, the accuracy, completeness, and origin of training data – specifically, who collected it, how and why, can directly influence predictions and subsequent decisions. To avoid data biases, and not to overlook marginalised or digitally invisible populations, dataset representativeness should be tracked and measured. Without this, AI tools risk reinforcing the needs of the majority during crises, while overlooking the specific needs of smaller or more vulnerable groups, defined by factors such as gender, region, religion, or ethnicity.

Reflecting the importance of data quality, especially if AI is increasingly used for crisis preparedness and response, European data spaces need to be complemented with clear and comprehensive data quality standards, checks and guidelines.

Box 7. Europe's critical dependencies on external data infrastructures for crisis management

An increasingly urgent concern, also stressed repeatedly by experts consulted for this evidence brief, is the EU's structural dependence on data infrastructures and publicly accessible databases controlled by actors outside Europe (Mallapaty, 2025). Recent disruptions to widely used information systems have exposed the fragility of current arrangements and demonstrated the degree of dependence on external providers.

The temporary but near-complete shutdown of the Famine Early Warning Systems Network (FEWS NET[35]) in early 2025 following US federal budget cuts eliminated, for nearly a year, one of the two main pillars of global famine early warning. FEWS NET had provided critical food security forecasts for more than 30 vulnerable countries for nearly four decades, and its abrupt closure left humanitarian

34 https://centre.humdata.org/quality-measures-for-humanitarian-data/

35 https://fews.net/

actors, including European organisations operating under DG ECHO funding, without early warning data, despite escalating food crises[36]. Similarly, the National Oceanic and Atmospheric Administration (NOAA[37]) has cuts that make the availability of weather and climate data used globally for disaster preparedness and response uncertain. Similarly, funding cuts to WHO and staff lay-offs at Centers for Disease Control and Prevention (CDC) (especially the data departments), may limit the availability of epidemiological and health data[38].

Beyond these cases of government-funded infrastructure, the 2023 closure of Twitter's (now X) free Academic API and its transition to commercial tiers severely disrupted social research (Blakey, 2024). It included crisis mapping and real-time social media analytics, which affected researchers and humanitarian organisations who had relied on Twitter-based situation monitoring and analytics. Starlink had been extensively used in the initial months of the war in Ukraine, but business or private interests then led to changes in the availability of services, even though several states paid for the service to be maintained (Abels, 2024). The outage of Amazon's web services in October 2025 (see Figure 10) further highlights the vulnerability of web services, apps and platforms on privately-owned infrastructures[39]. This leads to questions of how much of critical AI, web and communication infrastructures should be left to private companies, especially if these companies are based outside Europe.

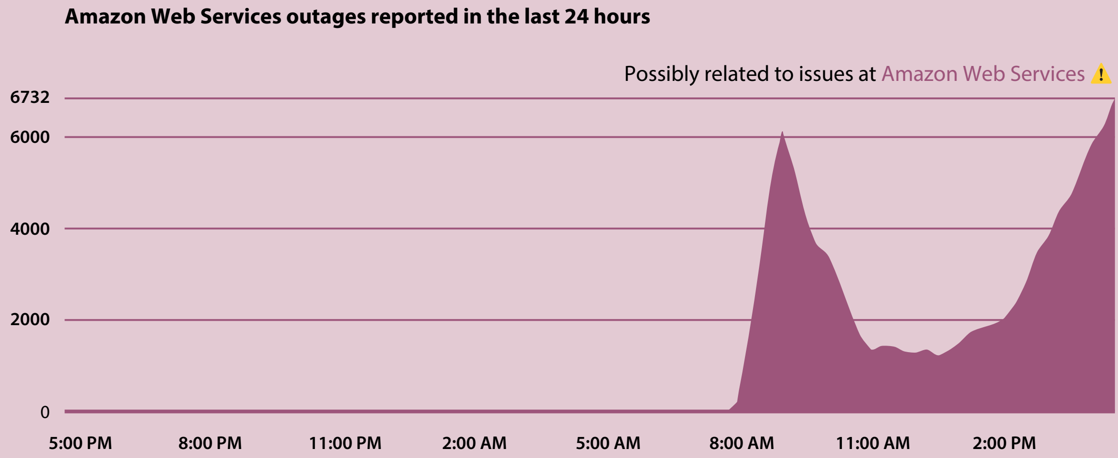


Figure 10. Down detector© tracking of Amazon web services outages.
Retrieved October 20, 2025, from https://downdetector.com/status/aws-amazon-web-services/

Ensuring data are collected and stored in a standardised way, and that shared data are interoperable and of high quality is an ongoing challenge. Poor data governance can lead to AI systems drawing the wrong conclusions; for example, if an earthquake damage model is trained only on data from regions

36 https://www.thenewhumanitarian.org/analysis/2025/03/25/humanitarian-data-drought-deeper-damage-wrought-us-aid-cuts, https://icha.net/2025/06/16/fews-net-returns-but-can-we-still-rely-on-it-for-famine-early-warning/

37 https://www.noaa.gov/

38 https://www.nature.com/articles/d41586-025-03365-1

39 https://www.theguardian.com/technology/2025/oct/20/amazon-web-services-aws-outage-hits-dozens-websites-apps

with certain building types, it may mis-predict impacts elsewhere. Importantly, AI cannot compensate for a lack of robust, standardised, and interoperable data. Addressing challenges of data access, quality, and availability gaps is an essential precondition for the reliable and trustworthy use of AI in emergency preparedness and crisis response. Therefore, standards for data collection, formatting, metadata, and validation must be agreed upon at EU level.

Privacy and legal compliance. Crisis-related data often include personal information (for example, mobile phone location data of people in an affected area, health records in a pandemic, surveillance footage used to find victims). Sharing such data across agencies or with private partners must respect GDPR and other privacy laws. Governance solutions include data anonymisation or pseudonymisation techniques, data-sharing agreements with strict purpose limitations, and real-time oversight by data protection officers during emergencies. GDPR does allow flexibility in disasters (processing data to save lives is permissible)[40], but this is not a carte blanche; it must be necessary and proportionate[41]. An ethical data governance challenge is to balance urgent lifesaving use of data with individuals' rights. Clear guidelines (and possibly pre-approved emergency protocols) are needed so that responders know how to share data legally (for instance, between a telecom provider and a civil protection authority, to identify concentrations of survivors), without delay or fear of later legal repercussions. The 'data responsibility' approach advocated in humanitarian operations calls for doing the maximum good with data while minimising harm, through privacy impact assessments, even in crises.

Cross-border and inter-agency coordination. Disasters do not respect borders, and under the EU Civil Protection Mechanism, assistance and information often flow between countries. A governance challenge is aligning data governance and AI practices across different jurisdictions and organisations. For example, as already stated above, the GDPR follows a comparable logic; it applies to (1) entities established in the EU, regardless of where data processing occurs, and (2) non-EU entities that offer goods or services to, or monitor the behaviour of individuals in the EU, including during crisis contexts. Thus, both EU and non-EU actors using personal data in disasters affecting EU citizens or occurring within EU territory are subject to GDPR obligations, including a lawful basis for data processing, data minimisation, and safeguarding measures. In short, during disasters, both the AI Act and GDPR apply based on the location of the affected data subjects and/or AI system use, not merely the location of the technology provider. This jurisdictional scope ensures that core principles of data protection, fundamental rights, and AI accountability remain in force, even in urgent cross-border disaster response operations.

Some countries or agencies may have more advanced AI capabilities than others, and data policies may differ. The Emergency Response Coordination Centre (ERCC) and the newly established Union Civil Protection Knowledge Network are platforms that can promote common standards. For example, if multiple countries are using AI-based early warning systems, coordinating their alerts and criteria becomes important so that an EU-wide overview (a situational picture) is consistent. International

40 Recital 46 - Vital interests of the data subject.

41 Information Commissioner's Office. (2024). Data sharing: A code of practice, UK GDPR guidance and resources. Retrieved from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/data-sharing-a-code-of-practice/data-sharing-in-an-urgent-situation-or-in-an-emergency/

public–private data sharing (e.g. with global platforms or satellite providers) requires negotiating terms that comply with EU standards (such as not violating EU citizens' privacy), an area that often requires political agreement and trust-building.

Until the AI Act fully comes into effect (by 2026), there is a grey zone where actors rely on general principles (Gstrein, Haleem & Zwitter, 2024). Uncertainty about liability, or on what counts as acceptable 'experimentation' in emergencies, can slow down adoption. For instance, emergency services might hesitate to use an AI tool for real-time decision support without clear approval or standards, fearing what might happen if it failed (Comes, 2024). The European Commission's scientific advice bodies and regulators need to provide practical guidelines, codes of conduct, and perhaps sandbox environments for AI in disaster management. This would allow the testing of innovative AI under supervision and with ethical oversight – similar to a controlled 'pilot' – before scaling up to full deployment.

An emerging idea is to incorporate continuous risk assessment and 'red-teaming' (see footnote 24) for critical AI systems, even after deployment, to catch problems early. Ultimately, as Dr Daniela Mahl remarked about AI in health emergencies, *'AI's ability to support or undermine public health efforts depends on how it is governed and implemented. The line between innovation and harm is thin, especially in high-stakes emergencies'.* [42] This underlines that robust governance – through clear rules, inter-sector collaboration, and oversight bodies – is as important as the technology itself (Chun, Octavianti, Dogulu, & Tyralis, 2025).

42 World Health Organization. (2025, May 23). *Responsible AI use can advance risk communication and infodemic management in emergencies, new study shows* [Press release].

6. Data preparedness

Introduction

The performance of AI models crucially depends on the availability of large amounts of high-quality data to train and calibrate the models before a crisis occurs, and to tailor models to the context and situation at hand. As such, the performance of AI in crisis management depends on the availability and integration of data sources across the different phases of crisis management. However, data-sharing protocols, data availability and data standards may be very heterogeneous, particularly in cross-border or cross-national operations. As such, data may arrive late, under unclear legal bases, or in incompatible formats (Raymond & Al Achkar, 2016). Operationally, field research has shown that a lack of pre-negotiated agreements and role-based access is a primary bottleneck for efficient crisis response (Comes et al., 2020; Turoff, Chumer, de Walle, & Yao, 2004; Van de Walle & Comes, 2015).

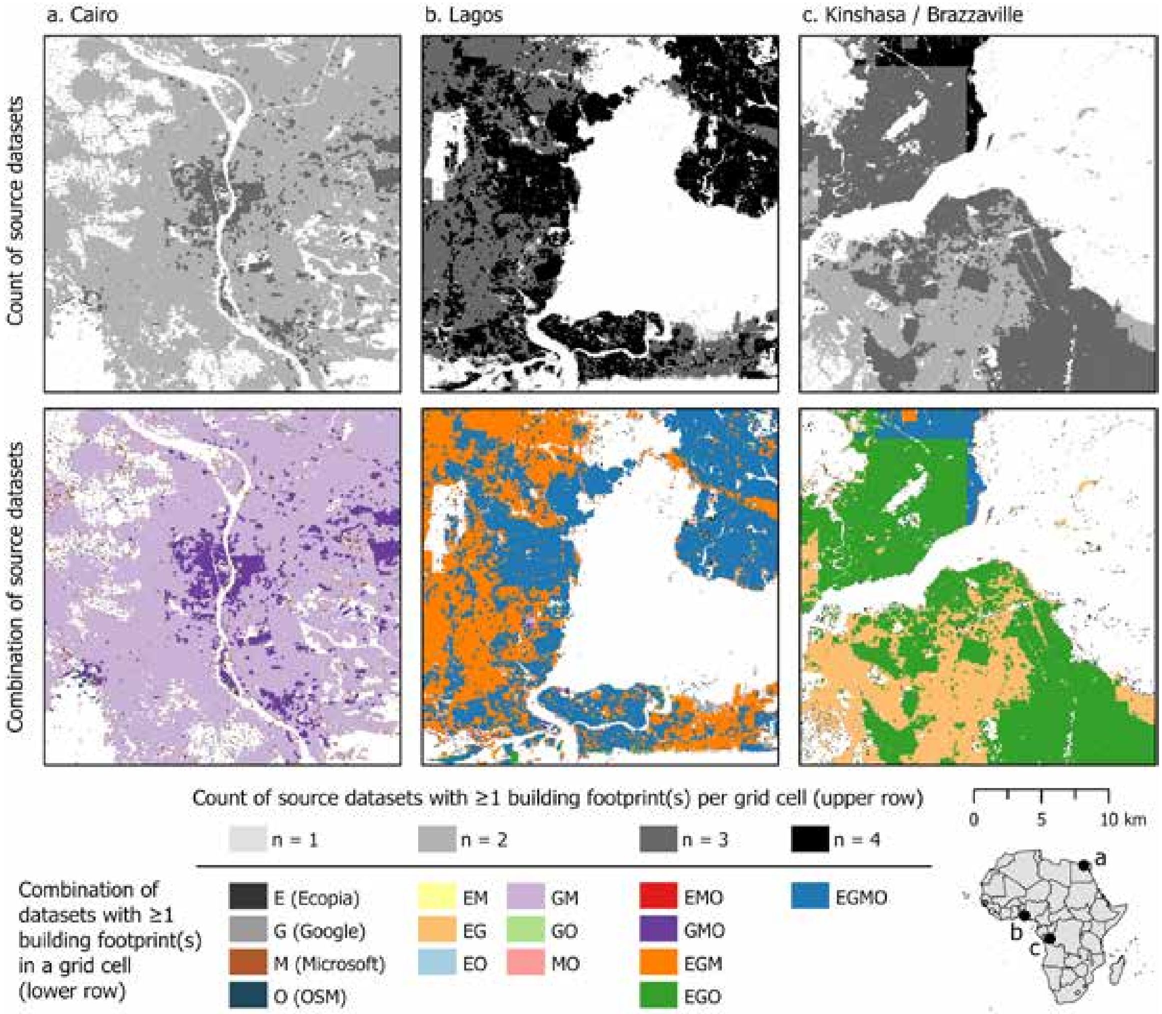


Figure 11. *Comparison of building footprints for three different cities* (from Chamberlain et al., 2024).

The *SAPEA Evidence Review Report on Strategic crisis management in Europe* therefore frames 'data-preparedness' as clarity on what data are needed for which decisions, who controls them, how they can lawfully be shared, and what quality and interoperability thresholds authorise their use (SAPEA, 2022). Roadmaps for responsible AI for crisis similarly stress data governance, provenance, representativeness and uncertainty disclosure as prerequisites for trustworthy AI (Lee et al., 2022).

At the same time, AI offers new opportunities, especially for data sparse contexts. While the initial assumption may be that such contexts do not allow for the use of AI, increasingly, AI is used to bridge and fill in possible blanks, for instance by using Deep Learning to map the built environment (Gevaert et al., 2024). While these advances are promising, different models and tools still show stark differences (Chamberlain et al., 2024), as highlighted in Figure 11 for the examples of Cairo, Lagos and Kinshasa.

Considerations for data preparedness

Access and legality. Support for humanitarian data preparedness and data responsibility include guidelines and operational practices for risk-benefit assessment, risk minimisation and incident response (Raymond & Al Achkar, 2016; Raymond et al., 2016).

Data quality and fitness-for-use. Data-quality dimensions such as accuracy, completeness, timeliness, consistency, interpretability clearly affect outcomes of data-driven decisions (Aylett-Bullock et al., 2022; Ebener, Castro, & Dimailig, 2014). While FAIR principles, normally used for scientific research, recommend that data be Findable, Accessible, Interoperable and Reusable, with rich metadata and persistent identifiers (Wilkinson et al., 2016), there are specific guidelines for disaster data and information management put forward by UN OCHA that should be merged with these, upholding principles like reciprocity and humanity (Van de Walle & Comes, 2015). Documentation, such as datasheets for datasets and model cards, make scope conditions explicit and auditable (Mitchell et al., 2019).

Availability, robustness and resilience. Crises destroy infrastructure, including telecommunication and digital infrastructure. Consequently, data preparedness must have back-ups and be designed to tolerate outages and limited bandwidth. Technology such as satellite-supported telecommunications have been used in Ukraine especially (Abels, 2024) and may also be promising for other crises. Evidence from crisis information management stresses the need for analogue fallbacks (Mendonça et al., 2007) and graceful degradation to human control (i.e. the ability for automated systems to reduce their functionality and transfer control back to human operators) (Edwards & Lee, 2017). The literature on meaningful human control (Cavalcante et al., 2023) also suggests the notion of *shared control* to balance control ability and authority, especially in rapidly changing and complex situations.

Interoperability. Europe has powerful data infrastructures, such as the DRMKC Risk Data Hub for risk/loss data; GDACS and Copernicus and the INSPIRE database for spatial information. However, interoperability across the different countries and regions, and integration with national data systems remain a challenge (SAPEA, 2022).

Towards a data preparedness framework

As emphasised, AI depends on high-quality data. Common causes for gappy, uncertain and incomplete data are access that is limited or too late (the lack of data-sharing agreements during response), context misfit (data are not representative for the context of the response, the model was trained for another context), a lack of interoperability (data cannot be integrated), and unclear data origin (data cannot be verified and are not trustworthy). These mechanisms explain why AI may either be rejected or used inappropriately. It is important to recognise that this is a socio-technical system, meaning that both the technical components and the human and social aspects of its use must be carefully considered. By following humanitarian response approaches (Aylett-Bullock et al., 2022; Raymond & Al Achkar, 2016; Raymond et al. 2016; Van Den Homberg, Visser, & Van Der Veen, 2017), and using data sharing agreements[43] which span many countries and organisations, data preparedness and the development of a dedicated pipeline can be helpful, but also adapted to EU regulations and law.

Complementary frameworks include UNDRR's new DELTA Resilience initiative[44], which is developing a Data Maturity Ecosystem Assessment that has been applied across several countries[45] and includes dimensions around data, infrastructure and tools, and also questions around institutional readiness and governance. These frameworks could provide indications of digital maturity across different contexts.

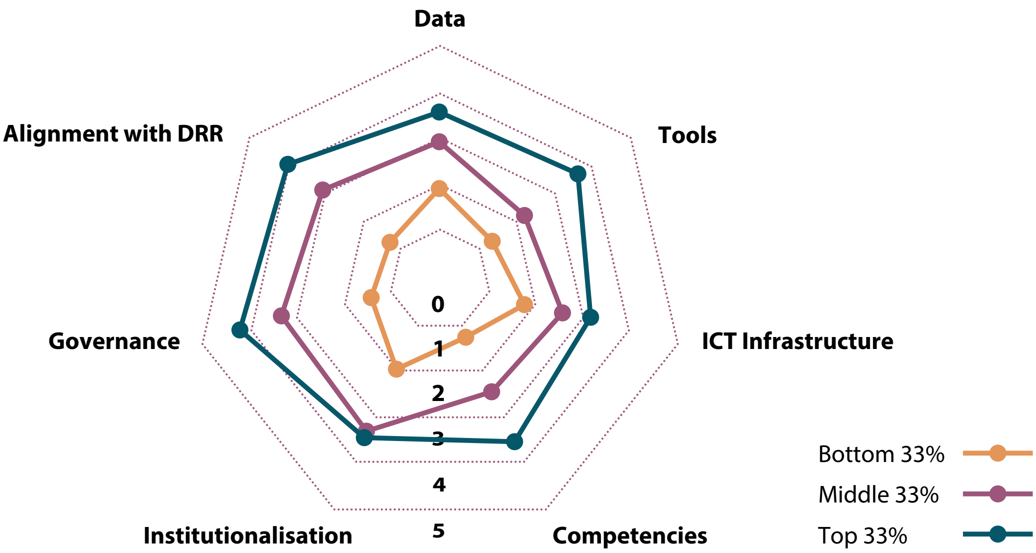


Figure 12. Result of UNDRR's digital and data maturity assessment for various countries, highlighting stark differences in maturity across countries. Retrieved from https://www.undrr.org/publication/documents-and-publications/data-and-digital-maturity-disaster-risk-reduction-informing

43 https://www.climatecentre.org/wp-content/uploads/Climate-Centre-NMHS-Guide.pdf

44 https://www.undrr.org/building-risk-knowledge/disaster-losses-and-damages-tracking-system-dts

45 https://www.undrr.org/media/84892/download?startDownload=20251016

In line with the previous *SAPEA Evidence Review Report* (SAPEA, 2022), the following framework may provide a way forward to strengthen the EU's Crisis Management Data Preparedness:

1. **Understanding and mapping information needs.** Identify high-value decisions, (AI) models and required datasets for training and response; specify acceptable errors, uncertainty and coverage.
2. **Pre-position data.** Acquire and prepare the best-available datasets and reference data (such as Common Operational Datasets) for crisis preparedness, to support key decisions. Identify the required granularity, quality availability of the datasets across Member States and determine sharing protocols for data that cannot be shared before a crisis.
3. **Establish data standards** to ensure data interoperability across Member States. This includes metadata, and agreements on data collection processes, completeness and coverage, as well as reporting frequencies before and during a crisis (when information flows need to be much more frequent).
4. **Training.** Ensure data literacy. Establish training and drills in managing information from different sources. Develop documentation.
5. **Learning.** Feed lessons back into knowledge systems and consistently build capacity.

The following table seeks to synthesise the different requirements for data preparedness by distinguishing the protocols and agreements that need to be in place (Step 2 in the above) from the data standards that need to be agreed (criteria based on data quality frameworks) in Step 3.

Data preparedness requirement	Relevance	Crisis management phase	Mechanisms
Data prepositioning protocols and agreements, for instance with respect to			
Protocols for lawful access	Enables timely, legitimate sharing across public/private data owners; seek to minimise potential data harms	Preparedness/Response	Pre-negotiated DSA; role/ attribute-based access; activation triggers; incident-handling and takedown (Aylett-Bullock et al., 2022)
Security and privacy by design	Protects sensitive data but ensures access if needed	All phases	Data minimisation; retention/expiry by default; access logging, role/attribute-based access
Data standards and data quality requirements, for instance			
Traceability and verifiability	Links outputs to inputs, enabling accountability and learning	All phases; critical post-incident	Versioned datasets; persistent identifiers; provenance logs
Accuracy, representativeness, coverage and equity	Reduces bias and uneven coverage that distort targeting and prioritisation	Preparedness/Recovery	Sampling audits; coverage maps; impact assessments; community review (Lee et al., 2022)
Relevance	Supports fit to local context	Preparedness/Response	Uncertainty quantification; local validation protocols (Mitchell et al., 2019)
Interoperability	Cross-border operations require shared schemas and IDs	Preparedness	Data quality and format alignment, clear and unique identifiers
Timeliness and robustness	Functions under disrupted communication and time pressure	Response	Prepare Common Operational Datasets (CODs); backups and caches; graceful degradation; analogue fallbacks (Van de Walle & Comes, 2015)

Table 3. Data preparedness requirements.

7. User uptake

Introduction

Across disaster management, AI is increasingly used to predict (Charlton-Perez et al., 2024) and monitor hazards (Jones et al., 2023), map out damages rapidly (Bhadauria, 2024) and/or communicate with the population (Ray, Merle, & Lane, 2025). While AI is viewed as a 'game changer', especially in the medical field (Haykal et al., 2025), the uptake by analysts, incident managers, and field teams remains inconsistent, and the evidence is relatively sparse.

One of the central mechanisms to determine uptake is widely considered to be trust (Choung et al., 2023), where trust is moderated via (perceived) usefulness. Trust in a technology is defined as *'the willingness to depend on and be vulnerable to an Information System in uncertain and risky environments'* (Bach, Khan, Hallock, Beltrão, & Sousa, 2024), such as crises. As such, trust is necessary for technology's use under pressure. However, when trust outstrips capability or context-of-validity, it becomes overconfidence and feeds automation bias. Explainability and transparency matter (van Leersum & Maathuis, 2025), but evidence indicates that these are instrumental to gaining trust, not substitutes for it.

As crisis response is a distributed effort, decisions are made by teams and networks of organisations, not by isolated individuals (Comes, 2024). Consequently, we also analyse trust, and the implications of human–AI teaming and organisational routines for the uptake of this technology.

Building trust in AI

Foundational work on trust in technology, and especially automation, shows that what we trust depends on a number of factors such as having accurate mental models of the system, the degree of uncertainty, and the system's operating envelope (Lee & See, 2004). A miscalibration of our levels of trust in technology can produce *disuse* (too little trust) or misuse and overreliance (too much trust) (Hoff & Bashir, 2015; Parasuraman & Riley, 1997). Empirical studies particularly document automation bias, the tendency of people to use shortcuts and *'to use automation as a heuristic replacement for vigilant information seeking and processing'* (Mosier & Skitka, 1999). Training, instructions and team arrangements can influence biases. Related factors include algorithm aversion[46] after errors and avoiding AI once it made a mistake. Dietvorst, Simmons and Massey (2015) therefore caution that trust does not necessarily increase with exposure to technology.

Recent studies on the performance of AI-supported decisions emphasise that the *'best teammate'* is not necessarily the most accurate AI model; assessing performance primarily based on indicators such

46 The tendency of individuals to distrust or avoid decisions made by algorithms.

as precision or recall may therefore not be sufficient. Complementarity with human competences, predictable behaviour, selective deferral (where the AI asks the human for a decision), and transparency on uncertainty all improve the performance of human-AI teams (Bansal, Nushi, Kamar, Horvitz, & Weld, 2021; Hemmer, Schemmer, Kühl, Vössing, & Satzger, 2025). Cognitive forcing functions (to encourage deliberative and analytical thinking) and structured prompts to reconsider AI outputs can lower inappropriate reliance, especially if they pre-empt possible human misconceptions (Buçinca, 2024; Buçinca et al., 2025).

Box 8. From AI to human-AI teaming

Recognising that in reality, humans and AI increasingly form teams, there are calls to consider AI and humans as a social-technical system, and to understand the effects of collaboration. This includes areas such as Machine Behaviour (Rahwan et al., 2019), Human-Centred AI (HCAI) (Ozmen Garibay et al., 2023) and Hybrid Intelligence (Akata et al., 2020). HCAI in particular has emerged as a prominent field advocating for explainable, fair, and transparent AI systems that keep humans 'in-the-loop' rather than replacing them (Shneiderman, 2020). This human-centred approach seeks to design AI that enhances rather than diminishes human agency.

The literature specific to crises and disasters points to two amplifiers that can mis-calibrate trust. Firstly, there is domain shift; models trained on one region, type of building stock, or hazard type (flood, earthquake, etc.) degrade when they are applied in other contexts. This phenomenon is well documented in satellite image damage assessment and in crisis informatics (Manzini, Perali, Tripathi, & Murphy, 2025; Van de Walle & Comes, 2015). Secondly, there is organisational coupling; information abundance without coordination leads to cognitive and moral overload and reduced decision quality. Teams require role clarity, shared views on uncertainty, and auditable decision pathways (Altay & Labonte, 2014; Comes et al., 2020; Turoff et al., 2004).

Understanding AI as part of a team, interaction with the AI system becomes a determinant of trust and may create machine authority bias (Kudina and de Boer, 2021). Research in cancer recognition suggests that conditions of uncertainty (that are also typical during crises) tend to amplify bias towards machine credibility (Nickel, Kudina, & van de Poel, 2022). Training and experience are crucial factors in empowering people to overcome this bias; research in the medical field has shown people with less targeted experience tend to agree with the suggestions of AI systems faster and more frequently than practitioners with more professional experience (Tschandal et al., 2020).

AI systems in crises can impact the lives and livelihoods of communities deeply. If AI is used, research suggests that a meaningful dialogue is needed with different user groups (such as responders, analysts, communities) to improve transparency and trust, by discussing what AI can and cannot do (Visave, 2025).

In summary, trust is vital to ensure technology uptake, and especially AI. Inherently, trust is linked to tools being *useful* to the task at hand. However, with increased trust, there is also the problem of overconfidence and overreliance that may lead, over time, to deskilling (Vallor, 2015), or to flawed decisions. Overconfidence does not arise from a shortage of explanations; it arises when users do not understand whether the model is adequate for the task at hand (or crisis), how uncertain or reliable the model output is, and if, when and how suggested actions can be adapted or reversed if needed.

Recognising that crises involve groups and teams that need to collaborate, AI should also be built to support team performance under uncertainty, and not solely with model accuracy. In the preparedness phase, the priority is to specify scope conditions *(when and where can the AI be used?)*, test for behaviour in other contexts, and to document failures. In the response phase, teams need confidence bands, cues *'when not to use'*, and explicit pathways by which to overrule the recommendations of an AI tool, or the choices that automated systems make if certain thresholds are exceeded that the model was not trained and tested for. In the recovery phase, accountability and fairness are critical, as models affect how vulnerable communities and groups that received prolonged assistance are identified. While there are guidelines such as Europe's Trustworthy-AI Guidelines[47] (see section above) that provide basic principles that can be considered for the design and use of AI, these principles require contextualisation, especially for crisis management. The framework presented in Table 4 below adapts some of the most important principles around trust to guidelines and mechanisms for crisis management. To this end, we follow the EU's Trustworthy AI Principles (see above) and contextualise them for crisis management.

AI principle for crisis management	Why?	Crisis management phase	Mechanism
Human agency and oversight: Meaningful human control under time pressure	Prevents overconfidence; ensures reversibility if stakes are high and information uncertain and incomplete	Preparedness/response/recovery	Clearly identify decision points; define explicit override criteria; maintain traceable decision logs; rehearse crisis escalation paths
Technical robustness and safety: Contextualisation or robustness to domain shift	Portability across hazards/regions is limited; avoid that models mislead	Preparedness/response	Context-of-use model cards that make the underlying assumptions transparent; local validation; rollback
Transparency: Operational transparency and clarity on uncertainty	Teams need actionable information that is immediately useful, not generic, explanations	Response/Recovery	Plain-language capability statements; uncertainty bands; 'when not to use'; provenance tracing
Diversity, non-discrimination: Fairness and equity in decision support	Automated AI/ decision support models may be biased	Recovery/Preparedness	Representativeness checks; de-biasing protocols; impact assessments; redress/appeal; stakeholder testing
Accountability: Accountability in multi-actor networks	Responsibility diffuses across organisations, especially if AI is involved	All phases	Audit trails; assurance clauses in contracts

47 https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1/language-en

Privacy and data governance: Data governance and privacy by design	Rapid data sharing must remain lawful and proportionate	All phases	Pre-approved protocols and rights including retention discussion. Harm-mitigation and takedown procedures
Social and environmental wellbeing: Do No Harm	Maintain crisis ethics of do no harm	Response and recovery	Participatory testing with (marginalised) communities; transparency high-stake events; systematic logging and reporting of energy and water consumption to ensure sustainable deployment

Table 4. Principles of trustworthy AI for crisis management.

8. Examples of areas of application and case studies

Introduction

The following case studies and areas of application provide a means of illustrating AI in action within a crisis management situation. The case studies follow the axes and boundaries of AI for crisis management, described in Section 2.

Case study 1. **Artificial intelligence for disinformation detection**

AI technologies increasingly shape both the creation and detection of disinformation, revealing a growing 'arms race' between malicious actors and defenders. This dual-use dynamic illustrates how the same technology can simultaneously undermine and protect information integrity, especially in crisis situations. Understanding this tension is essential for developing responsible and resilient AI applications in the information domain. AI-based disinformation detection builds on advances in natural language processing to support emergency services, platforms and citizens within rapidly evolving information environments. While traditional models such as BERT[48] are effective within predefined datasets, they struggle with novelty and dynamic discourse. Large language models (LLMs) such as GPT-4 offer broader generalisation and enhanced analytic capabilities, but raise new challenges around bias, governance, and sustainability. Complementary human-computer interaction (HCI) approaches aim to foster user reflection and resilience against manipulative content. Some tools for monitoring disinformation include Bellingcat's Online Investigation Toolkit, *CrowdTangle, accountanalysis, attribution.news,* and Google *Transparency Ads Database,* which help track and analyse online misinformation campaigns. More details can be found with the *EU DisinfoLab*[49]. The different dimensions affected are as follows:

Temporal aspects. Disinformation spreads rapidly in crises, requiring timely detection. Authorities face resource constraints, making (semi-)automated analysis essential (Kaufhold, Rupp, Reuter, & Habdank, 2020; Riebe, Kaufhold, & Reuter, 2021). Smaller models like BERT (see above) often fail with novel data (Bayer, Neiczer, Samsinger, Buchhold, & Reuter, 2024; Lucas et al., 2023), while LLMs generalise better but performance still declines with unforeseen events (Jiang et al., 2024). Retrieval-augmented generation (RAG) can mitigate this by integrating up-to-date knowledge, yet human oversight remains critical (Lai et al., 2022). Understanding how AI-driven detection tools respond to rapid disinformation

48 A language model introduced by Google in 2018.

49 https://www.disinfo.eu/resources/tools-to-monitor-disinformation/

spread is crucial, enabling policymakers to allocate resources effectively and ensure timely, automated interventions during crises.

Spatial aspects. Disinformation operates across platforms and modalities. Video-sharing platforms integrate video, audio, text, and interaction (Niu et al., 2023), while voice messages exploit emotional cues (El-Masri, Riedl, & Woolley, 2022). Local misinformation in emergencies differs from coordinated global campaigns, requiring both scalable and context-sensitive responses. Knowing how AI models track misinformation across platforms and modalities helps authorities design platform-specific, AI-supported countermeasures and context-sensitive monitoring.

Stakeholder aspects. Emergency responders depend on prioritisation tools, while developers face dataset and bias challenges. Platforms must balance detection with free expression, and users may encounter interventions such as nudges or credibility indicators (Hartwig, Sandler & Reuter, 2024; Roozenbeek, Culloty, & Suiter, 2023). At the same time, detection technologies can be misused by state or non-state actors for censorship and surveillance purposes (Kiritchenko, Nejadgholi, & Fraser, 2021). Awareness of how AI detection systems interact with users, platforms, and responders ensures that policies balance automation, free expression, and ethical use of technology.

Data ethics and governance. Bias remains a major risk, with models inheriting existing stereotypes around race, gender, and age (An, Huang, Lin, & Tai, 2024; Oketunji, Anas, & Saina, 2023; Hartmann et al., 2025). Dataset inconsistencies exacerbate the problem, as definitions vary across annotators (the people who describe and apply labels to data) and cultures (Laaksonen, 2023; Goyal, Kivlichan, Rosen, & Vasserman, 2022). Without transparent governance, automated detection risks the unjust suppression of discourse, or the disproportionate targeting of certain groups. Recognising biases in AI and governance gaps is essential to preventing discriminatory outputs or forms of unjust enforcement, when automated detection tools are deployed.

Socio-technical framing. Disinformation detection is embedded in socio-technical systems. Automated classifiers (Burnap & Williams, 2016; Kaliyar et al., 2019; Bäumler et al., 2025) are complemented by human-computer interventions that foster reflection and media literacy (Ayoub, Yang, & Zhou, 2021; Hartwig et al., 2024). Transparent explanations increase trust (Kirchner & Reuter, 2020), while opaque designs risk public reaction (Wood & Porter, 2019; Nyhan & Reifler, 2010). Information cues based on indicators of provenance, such as propaganda techniques or source context, provide comprehensible alternatives (Sherman, Stokes, & Redmiles, 2021). Understanding how AI classifiers and human-computer interaction combine allows policymakers to promote trust, literacy, and responsible AI system design.

Performance boundaries. Model performance depends heavily on prompt design (Zamfirescu-Pereira, Wong, Hartmann, & Yang, 2023; He et al., 2024). Failures with out-of-distribution data and trade-offs in interpretability limit trust in high-stakes contexts (Nirav Shah & Ganatra, 2022). Hybrid approaches that combine automation with human judgement remain essential. Policymakers need to know the limitations of AI models in order to implement safeguards, provide hybrid human-AI oversight, and fail-safes in high-stakes contexts.

Environmental impact. As stated, training and operating LLMs consumes vast computational resources, generating high energy use and carbon emissions. While smaller models such as BERT are lighter, frequent retraining for crisis adaptation adds cumulative costs. Sustainable governance should address environmental concerns by promoting efficient architectures and shared infrastructures. Awareness of the computational demands of AI systems guides policies towards sustainable, energy-efficient, and responsible deployment of automated detection tools.

Case study 2. **AI for weather forecasting and early warning systems**

AI weather and early-warning systems (EWS) increasingly combine meteorological foundation models with geospatial impact models to deliver high-resolution, low-latency guidance, ranging from minutes ('nowcasting') to seasons. Evaluations show that for some events, AI prediction can match or outperform physics-based baselines while running at far lower latency, making it attractive for operational warning chains. In crisis management, these systems feed anticipatory action (e.g. forecast-based financing) and multi-hazard early warning, shifting decisions to earlier in time, where protective actions are cheaper and more equitable. A framing of AI that is user-centric, causal, and responsible is essential so that warnings are trustworthy, understandable, and effective (see in-depth discussion in Reichstein et al., 2025).

Temporal. AI's strongest temporal contribution is anticipatory support for weather forecasting – issuing earlier, more frequent updates across timescales (Lam et al., 2023; Espeholt et al., 2022). This enables preparatory measures such as pre-positioning, targeted cash disbursement, and/or surge staffing before impact. In real time, rapid updates can refine hazard footprints for events such as flash floods, windstorms, or heatwaves. Retrospectively, AI supports attribution and model improvement. Beyond hazards, early-warning systems (EWS) need to integrate exposure and vulnerability layers (Figure 13) to forecast who and what will be affected; for example, predicting crop yield losses in drought or potential hospital overload during heatwaves. Communication channels then translate forecasts into timely, actionable advice[50]. Temporal value is greatest when impact messages arrive before critical thresholds, and when they update dynamically as new data flow in.

50 https://wmo.int/activities/common-alerting-protocol-cap

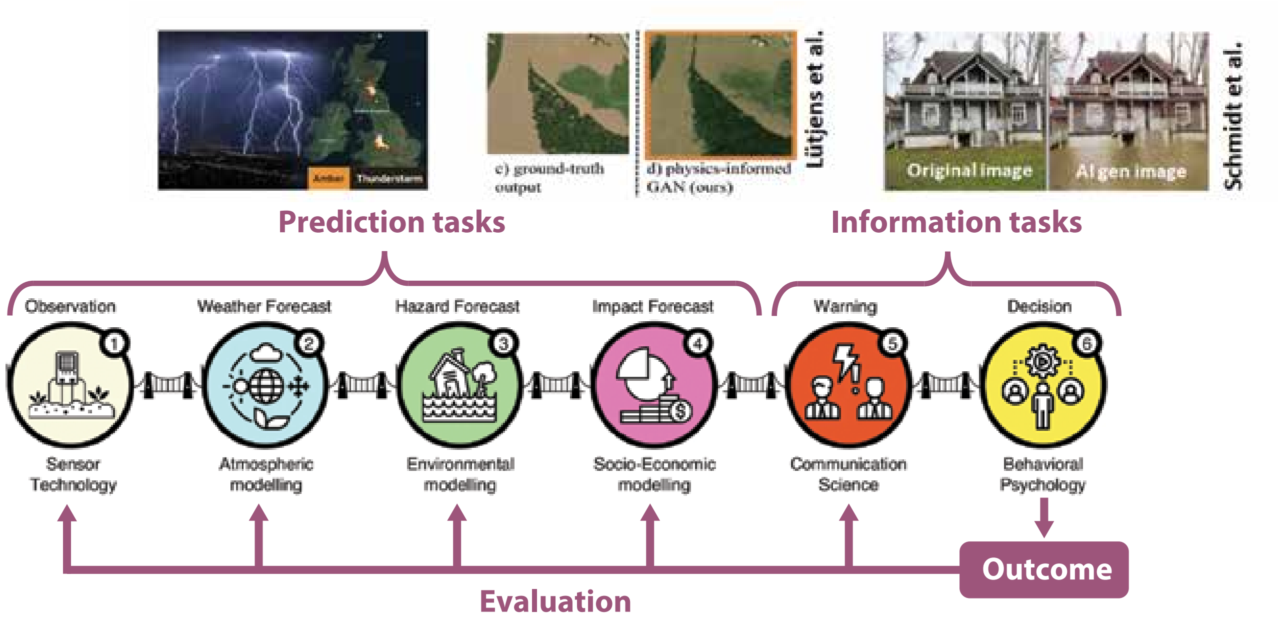


Figure 13. The Early Warning Chain and AI-enabled forecasting and communication tasks (from Reichstein et al., 2025).

Spatial. AI adds value to weather forecasting at both global and local scales (Mardani et al., 2025). Global models capture teleconnections and flow regimes; hyper-local downscaling translates signals into street- or basin-scale hazard footprints. This multi-scale capability supports cross-border coordination (e.g. storm surge monitoring across coastal states) while also maintaining municipal relevance. Impact models rely on high-resolution geospatial data (land cover, infrastructure, population) and local idiosyncrasies (levees, drainage). For instance, flood impacts differ sharply between urban districts with sealed surfaces and rural floodplains, but these can be learned with AI (Figure 13). Communication ensures that these spatially detailed products reach communities in formats that highlight local consequences and suggest protective actions.

Stakeholder. For weather forecasting, primary users are national meteorological and hydrological services, river-basin authorities, and civil protection agencies. Humanitarian actors plug forecasts into anticipatory pipelines; citizens usually interact with downstream warning products rather than raw model outputs. With impact forecasting and communication, the stakeholder scope broadens to utilities, health agencies, agriculture/food-security clusters, NGOs, media, and digital platforms (Tiggeloven et al., 2025). For example, heat-related health alerts may be co-produced with hospitals, while drought forecasts inform farming cooperatives. Co-production with emergency managers and communication experts is essential to calibrating thresholds, designing accessible products, and tailoring advice (e.g. the evacuation of clinics, closure of schools, protection of infrastructure).

Data ethics and governance. With weather forecasting, privacy risks are limited, since inputs are largely non-personal. Key issues include accountability when forecasts trigger costly actions; the explainability of black-box AI; lawful cross-border data sharing; and operational resilience (Kox et al., 2025). Provenance and uncertainty communication are essential. Impact forecasting and communication add critical governance layers, such as defining legal mandates across multi-agency chains, ensuring equitable coverage (especially in data-poor regions), and guarding against over- or under-warnings. For example, drought impact forecasts linked to food security decisions must be transparent and defensible. Evaluation should measure not only forecasting performance, but also actionability and harm reduction. Accessibility, multi-language delivery, and inclusivity for vulnerable groups are also highly relevant.

Socio-technical framing. AI weather forecasts are part of a human–machine system. Over-automation can degrade situational awareness; hybrid intelligence (i.e. co-activity, observability, directability) is the design goal (Alsamhi et al., 2024). Forecasters supervise and occasionally override AI; decision centres use AI-enhanced products alongside protocols and local knowledge. Impact forecasting and communication early-warning systems require closed-loop systems where monitoring, modelling, decisions and communication are tightly coupled. For example, wildfire spread forecasts must be translated rapidly into evacuation advice. Human-centred interfaces (e.g. salient cues, rationales, counterfactuals), graded protective advice, and tested message formats need to be researched further, including community and responder feedback loops.

Performance boundaries. AI weather forecasting performs best with repetitive hazards and dense sensing (e.g. European winter storms). Reliability declines under regime shifts or unprecedented extremes, where tacit knowledge is crucial (Sun et al., 2025). Human judgment remains essential to detecting drift, adjudicating conflicts and revising mental models. Impact models risk failing when exposure or vulnerability data are outdated, biased, or incomplete, or when compound risks dominate (e.g. heat + blackout + wildfire smoke). Communication may falter if trust is low or channels inaccessible. Stress-testing with scenarios, causal learning, and independent reviews can help bound use. Where uncertainty remains high, robust, low-regret (relatively low-cost/high-benefit) measures (e.g. opening cooling centres during heat alerts) and human override are key.

Communication and behaviour change. Hazard warnings typically reach people via authoritative bulletins or apps. Clarity, timeliness, and consistency of these messages determine protective behaviour, such as seeking shelter during severe thunderstorms (Scolobig et al., 2022). AI for impact forecasting and communication enables tailoring by channel, timing, and phrasing while guarding against inequity or manipulation. For example, localised flood alerts can specify which streets may be inundated. Best practice includes authoritative branding, plain-language and action-oriented advice, clear severity scales, probabilistic risk translated into concrete consequences, and accessible, multilingual formats. Research through field trials and post-event surveys are required.

Environmental impact. AI-based weather forecasts (inference) are less energy-intensive than their physical-numerical counterparts (Lam et al. 2023). For impact forecasts and communication, many potential inferences (e.g. insights, predictions, conclusions) can happen, for example, from user requests. Compact, distilled models (a technique for using smaller models) and edge inference (i.e. processing on local servers) reduce dependence on central servers, maintaining service during outages and minimising footprint.

Case study 3. **AI and situational awareness in disaster response**

As discussed above, AI systems can play an important role in the response and recovery phases of disaster response. Such systems have the potential to enhance the speed, accuracy, and coordination of disaster response efforts by combining information from multiple sources, by planning the deployment of key response assets (such as UAVs and response personnel) and by providing an ongoing visualisation of the unfolding situation.

To bring these possibilities to life, this case study focuses on a specific deployment (ORCHID[51]) to provide a degree of granularity on the actual use of AI technologies in the disaster response domain. Specifically, it provides an early example of one of the first real-world deployments of HABA/MABA systems for disaster response (being used during the 2015 Nepal earthquake).

Temporal aspects. The deployment in Nepal required AI systems to process and structure vast amounts of unverified, heterogeneous data (Ramchurn et al. 2016), arriving at different times over the course of the response. Crowdsourcing and machine learning techniques were employed to filter, classify, and geo-locate real-time reports, significantly reducing the cognitive burden on human operators. This allowed humanitarian responders to identify critical needs, such as blocked transport routes and affected population centres, with greater precision and timeliness than traditional methods alone could achieve (see Lorini et al., 2024 for a broader discussion of these issues). Rescue Global operators noted the system's value in accelerating their ability to deliver aid and allocate resources under highly uncertain and rapidly changing conditions.

Spatial aspects. This deployment was targeted at a sub-national level, on specific towns and cities that were impacted by the earthquake. Some of these areas were densely populated (e.g. Katmandu) and there was a good set of background data about the area. Other areas were more remote and distant and only sparsely represented in terms of their pre-existing knowledge. Being able to operate across these different levels of certainty and prior information were central to the success of the deployment.

Stakeholders. The system requires the close cooperation of multiple individuals and teams, from different organisations. It means there was not a single centralised AI system, and different stakeholders joined and left the operation during the unfolding response. Such systems are therefore well-suited to agent and multi-agent solutions, where each individual or organisation is represented by its own agent with its own resources.

Data ethics and governance. At the beginning of the response, the Rescue Global team only had access to publicly available data sources in the affected areas. Over time, this was augmented with additional information provided from a variety of official sources and by contributions from the first responders on the ground and members of the public who were situated in the impacted areas.

51 https://web-archive.southampton.ac.uk/www.orchid.ac.uk/

Socio-technical framing. Central to this system was the development and use of Augmented Bird-Table technology (Jones et al., 2015), which facilitated shared situational awareness across diverse teams. The ABT visualisation integrated heterogeneous data streams, including satellite imagery, crowdsourced information, and social media feeds (see United Nations Office for Disaster Risk Reduction & CIMA, 2024) through advanced AI-driven fusion and reasoning mechanisms. By doing so, it provided decision-makers with a unified, interpretable picture of evolving ground conditions (see (Kim & Boulanin, 2023 for a fuller discussion of this issue). Importantly, the system embodied the notion of Human-Agent Collectives (Jennings et al., 2014), in which humans and AI agents work together to analyse incoming data, prioritise tasks, and recommend actions in real time.

From a scientific perspective, the Nepal deployment provided strong validation for the concept of trusted human-agent collectives in high-stakes environments. The AI technologies deployed were not designed to replace human expertise but rather to augment it, ensuring transparency, adaptability, and resilience. The findings underscore the importance of robust AI-human collaboration models that respect the expertise of field operators, while leveraging computational advantages in data fusion, inference, and predictive analytics. It also highlighted that such responsibilities may need to change over the course of the response, depending on the workload of the human operators, and the operators need to build up trust in the AI system in order for it to function in an effective manner.

Performance boundaries. While this deployment helped save a number of lives in Nepal, many open issues remain to be solved. In particular, the underlying principles of what tasks and activities are best addressed by the human operators and what are best tackled by the AI system (Hernandez & Roberts, 2020), how AI systems can learn to be effective team members and good collaborators (in this system, the humans had to adapt their way of working to fit with the AI system), and how complex and evolving systems can be best visualised to enhance situational awareness (see Stauffer et al., 2023).

Communication and behaviour change. The deployed system led to a change in the way that Rescue Global responded to disasters. Their first responders interacted closely with the AI systems to generate, refine and test their plans. They used the AI systems to produce initial plans when they were under significant time pressure, rather than starting from scratch themselves. They then used the AI systems to monitor their plans and check if any of their underlying assumptions were violated. This task-sharing freed the humans from some of the detailed working and re-working, letting them focus on the effectiveness of their response and also re-plan in a more agile way, as they had the capacity and tools to try multiple scenarios.

In the longer term, the partnership between ORCHID researchers and Rescue Global demonstrated how AI can meaningfully contribute to humanitarian disaster response. By enabling shared situational awareness, accelerating data-driven decision-making, and supporting coordinated action, the deployment in Nepal stands as an important illustration of the operationalisation of AI for real-world crises.

Environmental impact. The system was deployed from a rapidly assembled control centre that was established in the disaster zone. There was limited existing infrastructure for communications and access to significant computational systems. The individual agents and their interactions were run from inter-connected laptops that were brought into the response hub. There was the ability to process images and data remotely as and when they came in, but the system and its operation were comparatively lightweight in terms of use of compute resources.

Case study 4. **The use of AI during the COVID-19 pandemic**

The use of AI systems during the COVID-19 pandemic was widespread. The aim was to optimise the efficacy of governmental responses and alleviate the burden on healthcare personnel. Additionally, computer vision and thermal imaging were used in cameras and drones to monitor large groups of people in public spaces and travel hubs (e.g. Barnawi, Chhikara, Tekchandani, Kumar, & Alzahrani, 2021; Ding, Shang, Xie, Xin, & Yu, 2025). This case study looks at the use of AI for (1) contact tracing (2) image recognition of chest scans for medical decision-support and resource allocation; and (3) enforcement of quarantine measures.

Temporal aspects. In the early stages of the pandemic, AI was implemented primarily to facilitate contact tracing and analyse the resulting data efficiently. The launch of the Google/Apple Exposure Notification Application Programming Interface (Google/Apple API) in April 2020 marked a major development that later underpinned many national contact-tracing applications. This interface enabled decentralised and privacy-sensitive reporting of COVID-19 exposure through a combination of Bluetooth technology and cryptography, which national initiatives often augmented with AI data mining for comprehensive data analysis and predictive modelling of the pandemic (e.g. Ahmed et al., 2020; Mbunge, 2020). AI was also implemented to speed up recognition of COVID-19 in *chest imaging* and other decision-support cases. Systematic review studies reported already in 2020 on the diagnostic and predictive value of over 700 AI models, geared specifically to the COVID-19 response (e.g. Wynants et al., 2020; Röösli, Rice, & Hernandez-Boussard, 2021/online first in August 2020), a testimony to the speed of technological development during the pandemic. AI was also used to facilitate *quarantine measures*, for example, via real-time image analysis of individuals' selfies, coupled with GPS location (e.g., Ding et al., 2025; Fan, Wang, Deng, Lv, & Wang, 2022; Lee and Kim, 2022; Lashkul, 2021; Brewczyńska, 2020; Tarkhanova, 2023).

Spatial aspects. In the case of *contact-tracing*, Google/Apple API used individual smartphones to store proximity data locally. Apps developed with this API incentivised individuals to share data voluntarily with epidemiologists and anonymously with other people (e.g. van Brakel, Kudina, Fonio, & Boersma, 2022; Anom, 2022; Yang, Heemsbergen, & Fordyce, 2021). Centralised contact tracing (e.g. in South Korea, China) collected data across a range of public and private settings, including smartphones, transportation systems, bank cards, and CCTV recordings (Yang, 2022; Fan et al., 2022). Another approach featured visited locations that required scanning a QR code and storing the data locally, while aggregating the responses at national level (e.g., New Zealand) (Yang et al., 2021). In cases of *quarantine (self)enforcement,* AI systems were used inside individuals' homes. Some systems utilised geofencing algorithms in conjunction with Bluetooth and Wi-Fi (e.g., Hong Kong), while others intermittently prompted users to verify their location through real-time selfies combined with GPS data (e.g., Poland [Brewczyńska, 2020] and Ukraine [Lashkul, 2021; Tarkhanova, 2023]).

Stakeholder aspects. In contact tracing, large tech corporations played a crucial role in facilitating pandemic monitoring and influencing public health policy, based on the digital infrastructure they provided (e.g., Google, Apple, Baidu, Alibaba, and Tencent). Governments and public health agencies were often receiving the data streams, and were then responsible for processing, integrating, and enforcing them. Such an integral corporate-government collaboration gave rise to significant public critique and mistrust (Sharon, 2021). It also resulted in concerns about dehumanising individuals as mere 'data points' to feed commercial interests and depriving them of democratic agency (Yang, 2022; Siffels, 2021). The *lung scan analysis* case involved the rapid development by researchers and national development agencies of image recognition and prediction models. The overwhelming burden on the healthcare setting contributed to rapid AI adoption. For instance, UK health workers reported that they had been forced to abandon any existing technology scepticism to embrace AI-based support, due to the overwhelming increase in work and a backlog of services (Nix, Onisiforou, & Painter, 2022). Staff shortages led to the engagement of clinicians without sufficient clinical experience (Ibid.), in a situation where experience is an important factor in challenging AI suggestions in decision-making (Tschandl et al., 2020). In the case of *quarantine enforcement*, public health authorities and law enforcement agencies played a key role in surveilling individuals' confinement. The restrictions on civil rights and liberties, as well as large-scale data collection practices, often required rapid legal amendments, resulting in public tensions between individual rights and liberties and public health solidarity in European contexts (Brewczyńska, 2020; Tarkhanova, 2023).

Data ethics and governance. In *contact tracing,* much emphasis was placed on privacy-preserving data mining and analysis. In some cases of decentralised tracing, public consultation suggested that the emphasis on privacy in contact tracing was deemed counterproductive, favouring instead solidarity-based data disclosure (Verbeek, 2020). Centralised models (e.g. South Korea, China, Australia) raised greater concerns about mass surveillance and *sphere transgression* (i.e. intrusion into societal domains). For instance, despite stripping data of individuals' names, South Korea reported multiple cases of the reidentification of data subjects. Dashboards tracking the timed location history of identified virus carriers also included their age, gender, and ethnicity, leading to successful social media initiatives by citizens to identify carriers through data triangulation. This resulted in mental and physical harm to the identified individuals and promoted COVID-19-related stigma in society (Yang, 2022). In cases of AI-facilitated *lung scans*, ethical problems stemmed from problem formulation bias, a lack of algorithmic transparency, and primarily, with data collection and provenance that, in post-COVID-19 validation, exhibited pervasive bias and amplified inequalities in healthcare (Wynants et al., 2020; Williams et al., 2020; Röösli et al., 2021; Leslie, Mazumder, Peppin, Wolters, & Hagerty, 2021; Delgado et al., 2022). Data were often collected from patients from a high socioeconomic background, leading AI models to perform poorly for underrepresented and vulnerable populations. Moreover, the data were often collected and combined inaccurately; for example, positive cases originated in one setting, while negatives cases came from another, preventing coherent data sampling. The rapid deployment of AI systems for diagnostic support often bypassed comprehensive validation, resulting in diminished effectiveness, uneven performance, and harm to certain populations. *Quarantine enforcement applications* necessitated legal adjustments. For instance, in both Poland and Ukraine, the mandatory use of the quarantine app required the disclosure of sensitive information to public health services (Brewczyńska, 2020; Lashkul, 2021; Tarkhanova, 2023). Poland's policy revealed important inconsistencies, for example, not mentioning law enforcement as a data disclosure purpose, while it was used precisely for this if the individual disobeyed the quarantine rules (Brewczyńska, 2020).

Socio-technical framing. In the case of the contact-tracing app in the Netherlands, when the technological system was deemed ready and safe (i.e. validated), the procedural approach (including parliamentary voting) resulted in its delayed introduction and marginal effectiveness. Citizens perceived the delay as a lack of government support for these systems (Kudina, 2021). Cultural attitudes played a key role in public acceptance and trust in quarantine *enforcement systems* (e.g., Ding et al., 2025), enabling sabotage and creative appropriation practices to bypass confinement. In the case of AI-based diagnostic support of chest scans, clinician-AI collaboration was imperative to achieving optimal diagnostic results (Nix et al., 2022; Leslie et al., 2021) but was often challenged by the disproportionate burden on clinicians. Factors such as fatigue, a lack of experience and digital skills could all contribute to the uncritical adoption of AI suggestions (Tschandl et al., 2020). This may be intensified by the design of the user interface, which could be overly suggestive in decision-making (e.g. through colour-coding, percentage emphasis, etc.) (Kudina and de Boer, 2021).

Performance boundaries. While app-based *contact tracing* promised a fast and accurate mass tracking of COVID-19's spread, its effectiveness proved to be questionable due to a low uptake in voluntary settings and high error rates. Tracing apps based on Bluetooth often produced a high number of false positives, causing people to stop using them. The need to maintain the GPS signal and the consequent rapid battery drainage in smartphones were among the top reasons why people chose not to use the voluntary contact-tracing apps. For AI-assisted *chest scans*, the vast majority of models, assessed in systematic reviews, had a high risk of bias (Wynants et al., 2020; Williams et al., 2020; Delgado et al., 2022). The main causes were datasets that were insufficient and disproportionally represented, data overfitting, and a lack of objective validation in favour of speedy adoption. *Quarantine enforcement apps* were prone to multiple errors related to a combination of facial recognition, WiFi access, and GPS location. Users reported multiple system delays, crashes, and too short a designated time to provide a selfie (e.g. only 15 minutes in the case of Ukraine's app) (Tan et al., 2020; Fan et al., 2022; Brewczyńska, 2020; Tarkhanova, 2023). This resulted in multiple false automatic reports of violating quarantine rules and deployment of in-person checks, thus contradicting the original narratives of cost-effectiveness and efficiency.

Environmental impact. There are no existing reports on the explicit environmental impact of AI systems during the COVID-19 pandemic. Since AI systems reviewed here mostly focused on image recognition, data mining, and analysis, it is possible to deduce that their training and use, just as outside of the COVID-19 context, required large amounts of energy and water, resulting in significant CO2 emissions.

Conclusions

The evidence reviewed in this report discusses the characteristics, opportunities and risks associated with the use of AI in crisis preparedness and response, and analyses how risks can be mitigated. AI offers the potential to improve crisis management capabilities, especially when it comes to processing large amounts of volatile and heterogeneous data, a key challenge in crisis management. By using AI, the prediction and monitoring of disasters can be improved; situational awareness and decision-making capabilities enhanced. At the same time, the deployment of AI requires careful monitoring to ensure compliance to legal and governance frameworks, avoid algorithmic biases and ensure meaningful human control.

AI excels at standardised, data-intensive tasks that are typical in frequently re-occurring disasters such as floods, wildfires or droughts. However, it is not yet equipped for interpreting different contexts, or for data-sparse and new situations, where no training data are available. The evidence shows that AI can effectively handle environmental monitoring, early-warning systems, damage assessment from satellite imagery, and social media processing. The evidence also suggests that AI demonstrates better performance in specific, well-defined crisis management tasks, particularly those that involve rapid processing and pattern recognition of large volumes of heterogeneous data. Machine learning approaches can be used to process satellite imagery, sensor network data, and social media feeds at scales that would be impossible for human analysts alone, supporting rapid damage assessment and situational awareness across multiple affected jurisdictions. AI is also good at repetitive tasks that may fatigue humans, such as continuous environmental monitoring, which is important for early-warning systems for floods, droughts, wildfires etc. Experimental uses also point to the potential of AI to increase human capacity limitations – e.g. in handling a surge of requests via chatbots.

However, performance can degrade when AI systems trained in one context are applied elsewhere. AI is not yet equipped to handle unprecedented situations for which there is no training data. AI trained on historical data from specific regions or hazard types do not perform well when applied to different contexts, a phenomenon known as 'domain shift' that is particularly relevant in the context of climate change where, for instance, wildfires may occur further north than would be typical historically. Here, there is an emergent body of work on transfer learning that, for instance, has shown promise in analysing social media data for changing crisis management contexts (Kejriwal & Zhou, 2020).

The use of AI for moral decisions is contested, since AI does not have a mandate to make these important trade-offs. Similarly, AI is not yet designed to understand the political coordination and institutional complexities inherent in UCPM[52] operations. Legal frameworks create opportunities for European AI; the EU AI Act's 'high-risk' classification for crisis management systems ensures that the necessary safeguards are put in place. Of course, this limits the applicability of tools that may be deployed elsewhere, but may then jeopardise the privacy or safety of citizens. Consequently, this

52 Union Civil Protection Mechanism

provides room and opportunity for developing European AI tools. Importantly, the GDPR emergency exceptions allow data processing to save lives, while maintaining privacy protections.

Data is essential for AI, and data governance is a fundamental part of AI for crisis management. Four possible data issues may undermine the effectiveness of AI: (1) late access to data, due to no agreements being in place before a crisis (2) context misfit, where models trained on data from elsewhere do not represent local conditions (3) poor interoperability between data across different national systems, and (4) unclear data provenance (undermining trust, credibility and transparency), and/or poor data quality. While EU infrastructures (such as the DRMKC, Copernicus, INSPIRE) provide strong foundations, harmonised standards and pre-positioned datasets are important.

Trust can determine the use of AI. Crisis managers can under-trust AI assistance or over-trust potentially flawed outputs. The evidence shows that effective systems require transparency (for example, providing uncertainty indicators), clear operational boundaries and areas of application, explanations, and explicit guidance on when not to use them. Cross-border coordination faces additional barriers from differences in national AI policies, data protection interpretations, and technical standards.

Policy options

The Report puts forward a range of evidence-based options for policy, which might support the EU in developing the use of AI in crisis preparedness and response, while mitigating risks associated with AI systems. We include the *'what, why and how'* for each policy option, alongside potential advantages and disadvantages of each option. Importantly, these policy options are presented as a catalogue of what *could* be done, with the associated advantages and disadvantages designed to facilitate discussion and prioritisation on the possible ways forward.

Option 1. Establish a European Crisis Management Data Preparedness Framework

What: Establish European common data standards, pre-approved data sharing protocols and mechanisms that are needed for preparedness and response, whilst preserving privacy and assuring data quality. Data heterogeneity is also important, by including data from the social sciences.

Why: As set out in the section on data preparedness, AI tools for crisis management critically depend on high-quality data. However, there remain issues with data gaps, lack of interoperability, and the need for data harmonisation across Member States. Furthermore, there is a lack of protocols to guide decisions on which data need to be shared before a crisis, or when a crisis occurs.

How: Extend existing EU data infrastructures (DRMKC, Copernicus, INSPIRE) with crisis-specific data-sharing standards and data preparedness protocols across Member States that respect privacy and the EU's data sharing standards, as well as the AI Act. Combine specific domain mechanisms that already exist (e.g. for health and pandemics via the ECDC) with dedicated crisis management expertise for different contexts and scenarios. The starting point can be pilots for some of the most frequently occurring or severe hazards in Europe that require cross-border collaboration, such as floods, heatwaves, or wildfires.

Advantages

- Ensures that data are interoperable across Member States and tools that are used
- Ensures that data can be shared when and as needed, based on pre-existing standards and protocols that are agreed upon by all Member States *a priori* – avoiding potential delays in case of an emergency, and ensuring compliance with the EU AI Act and the GDPR
- Is a prerequisite for training of European-wide AI for different hazards that can fit *all* relevant EU contexts, instead of national models.

Disadvantages

- Will require frequent updates as technology and legislation evolve
- Requires contributions and buy-in of all Member States

- The underlying data may be provided in different languages and formats (e.g. different place names or postcode/zip code formats) that may require translation and contextualisation (although could be assisted by AI in future).

Option 2. Provide AI literacy and training for crisis management

What: Dedicated AI literacy training for crisis management authorities, analysts and policymakers, with training around the different uses of AI for preparedness and response needed.

Why: As indicated in the section on trust, AI remains a 'black box' for many users. At the same time, AI is becoming a crucial tool for crisis management, as examples throughout this Report have highlighted. A lack of in-depth understanding of AI for crisis management can lead to both overconfidence and under-use of the technology. Moreover, during the pressure of crises, technology that users feel uncomfortable with tends to be discarded. To ensure that AI tools are used adequately, competently and efficiently, and that trust is adequately calibrated, AI literacy training for different AI users and uses is needed.

How: Integrate AI literacy into existing UCPM training programmes and emergency exercise protocols for responders to ensure that AI can be used, even under the time pressures and cognitive load of a crisis. Also, provide dedicated training for analysts and policymakers that addresses the use of technology (what it is; how to use it adequately; practices in prompt engineering), as well as how to use it for decision-making and situational awareness, avoiding possible bias, or trust and overconfidence issues.

Advantages

- Training can be used to mainstream and harmonise AI use and standards across Member States (see Option 1)
- Insights from training can feed back into the further development of AI
- Training can be used to build confidence and trust in the technology, as well as developing an understanding of the potential risks, so as to ensure that AI is used adequately and competently
- Training can also ensure familiarity with the different AI tools and their functionalities.

Disadvantages

- Training programmes need to be updated and extended
- With the rapid evolution of AI technology, training needs to be frequently updated.

Option 3. Develop dedicated AI evaluation frameworks and establish knowledge-sharing platforms

What: Develop clear benchmarks and evaluation protocols for the use of AI in crisis management that are aligned with the EU''s standards and guidance on AI. The evaluation of results, experiences and lessons learned can feed into a dedicated European AI knowledge- sharing platform.

Why: AI has gained increasing importance for crisis management. This Report echoes the need for better evaluation of both the use and uses of AI under different scenarios and contexts. Current benchmarks, especially for foundation models, focus primarily on generic problems (Reuel et al., 2024). However, insights from crisis management are still relatively sparse, even though the conditions of a crisis (high stakes, time pressure etc.) have shown to alter human sensemaking and decision-making. Yet, there is no dedicated benchmarking system. This is especially problematic for areas where there is no ground truth (such as resilience assessment), or where AI is already so ubiquitous that no counterfactuals exist anymore (e.g. processing of satellite imagery). The definition and introduction of AI for different benchmark problems should be evaluated via dedicated protocols and compared with other models or tools (e.g. physics-based models and simulations; human assessment) and include responsible data standards. To that end, standardised metrics and protocols are needed by which to evaluate AI in crisis management that also reflect the different functions and uses of the AI.

How: Initial crisis benchmark cases can combine guidance on Trustworthy AI that the EU already has in place with insights and requirements from the crisis management domain (e.g. as laid out in the *SAPEA Evidence Review Report*, 2022). The guidance must be operationalised to develop clear benchmarks, see Figure 14 below. For validation of the framework, start with cases for the EU where there are frequently re-occurring scenarios that require collaboration and where the technology is already very mature, such as early-warning and damage assessment. The results can then be combined with dedicated knowledge-sharing platforms to capture what works and what fails, to ensure cross-European learning (and potential implementation in training, see Option 2).

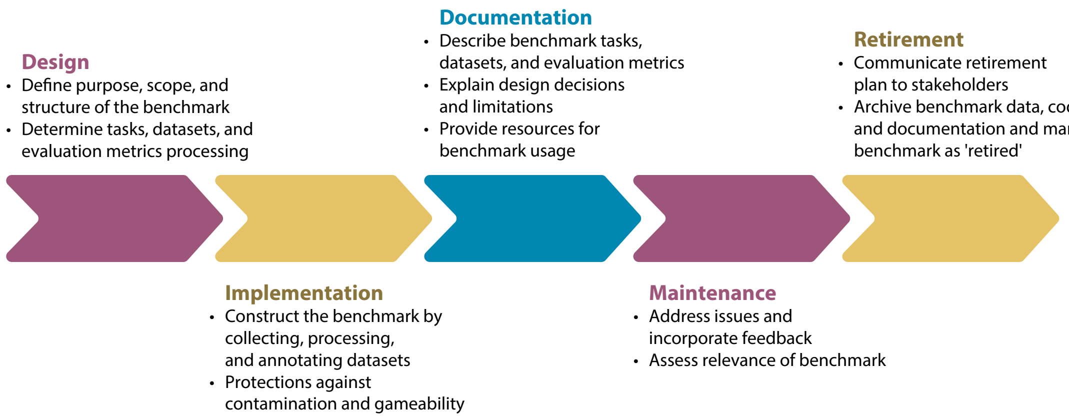


Figure 14. AI Benchmarking process (from Reuel et al., 2024).

Advantages

- Clear standards and benchmarks for the evaluation of different AI tools, to compare against physics-based, simulation models and/or expert assessment, and ensure that the EU's data standards can be maintained
- Such standards facilitate the comparison of different AI tools for different use cases and should facilitate procurement

- The standards are operationalised for the context of crisis management, making them more tangible and better contextualised than the higher-level principles that are already in place
- The benchmarks can also help track, compare and guide the development of new AI tools

Disadvantages

- Standards and benchmarks will need to be developed for different crisis scenarios and functionalities of the AI. This would mean, for example, acknowledging that AI for flood forecasting is different from chatbots for crisis communication or LLMs for reporting. This variety of functionalities will likely lead to a catalogue of benchmarks
- Given the rapid pace of technology developments, standards will need to be updated regularly to avoid too quick saturation or contamination
- Lessons learnt will need to be integrated regularly into training (see Option 2) and subsequent revisions of the benchmarks.

Option 4. Build European strategic autonomy for Crisis AI

What: Leverage the opportunities for dedicated AI for crisis management by strategically prioritising European AI.

Why: In the section on data preparedness, we discussed that many datasets, infrastructures, AI algorithms and technologies are currently developed and procured outside of the EU. These dependencies create vulnerabilities in data governance, trust, and system reliability. This is particularly problematic when crisis management involves sensitive information and requires accountability and adherence to EU standards. Furthermore, particularly for LLMs, the EU needs to acknowledge that the models are trained based on user input. Current dependencies may lead to vendor lock-ins and limited control over standards. To understand vulnerabilities, existing dependencies with respect to data, algorithms, and infrastructures and/or platforms (e.g. cloud services) need to be mapped out and prioritised according to their risks. To strengthen independence further, the coordinated procurement of EU-based AI with a specific focus on crises could be a way ahead to ensuring that requirements for European data residency and algorithmic transparency in crises are in place.

How: Map out data, algorithmic and infrastructural dependencies and vulnerabilities. Establish common procurement guidelines and standards requiring EU-based AI providers or European partnerships; invest in European AI for crises to address algorithmic dependencies; establish the necessary data infrastructure, standards (see Option 1) and backbone to train and develop AI, based on existing infrastructure, e.g. Copernicus or INSPIRE, or flagship projects such as the digital twin Destination Earth. Mandate that crisis management AI systems operate under European operational control.

Advantages

- In what is a critical sector, reduce vulnerabilities and strategic dependencies for data, algorithms, tools and infrastructure
- Foster European AI eco-systems
- Ensure transparency and that clear standards (for example, privacy and accountability), can always be upheld. Benchmarks (Option 3) may support this process.

Disadvantages

- As many tools are currently developed and deployed outside of the EU, a more limited selection of tools may be available and accessible at the start
- Potentially large-scale investment in data, infrastructure, and AI development is needed
- There may already be pre-existing contracts and dependencies that are hard to overcome, given the tendency for vendor lock-ins. Here, dedicated pathways and trajectories are needed to reduce dependencies over time.

Option 5. Ensure full compliance with the AI Act and GDPR in crisis contexts

Why: The AI Act and GDPR remain applicable during disasters. The AI Act applies to any AI system used or deployed in the EU, regardless of where it was developed. GDPR applies to personal data processing involving individuals in the EU or by EU entities, including in humanitarian crises.

How: Promote harmonised, crisis-adapted compliance protocols, e.g. streamlined risk assessments, predefined Data Protection Impact Assessments (DPIAs) for common AI use cases, and emergency data-sharing templates to reduce friction, while upholding legal standards.

Advantages

- Guarantees continuity of rights protection, even under crisis conditions
- Aligns with EU principles of trustworthy AI, reinforcing accountability and public trust
- Sets a global standard and promotes legal clarity for international actors

Disadvantages

- Can create operational delays during emergencies, due to documentation and oversight requirements
- May be difficult to enforce in non-EU jurisdictions or in fast-moving, cross-border disaster settings.

Option 6. Clarify legal responsibilities in cross-border and public-private operations

Why: Both EU and non-EU companies must comply with EU laws when operating in or targeting the EU. However, legal responsibility is fragmented within multi-actor environments, especially in public-private partnerships (PPPs) and non-EU humanitarian operations.

How: Develop an EU crisis AI governance framework for public-private partnerships, with model clauses on liability, data sharing, and ethical compliance, aligned with GDPR, the AI Act, and humanitarian data principles.

Advantages

- Clarifies liability and jurisdictional scope and ensures accountability across the AI lifecycle
- Reduces regulatory uncertainty for private partners and enables more robust contractual arrangements.

Disadvantages

- Requires complex coordination, particularly when multiple regulatory regimes overlap (e.g. local data laws in non-EU countries)
- May deter private actors from participating in crisis innovation without legal safeguards.

Option 7. Address legal and ethical gaps in non-EU humanitarian operations

Why: EU law does not always apply to the personal data of non-EU citizens in third countries. Yet EU-funded or EU-operated AI systems overseas raise ethical obligations, especially where vulnerable populations are involved.

How: Adopt binding data responsibility standards (e.g. Red Cross/OCHA frameworks) for EU-funded crisis AI activities outside the EU. Encourage 'ethics by default', even where GDPR does not formally apply.

Advantages

- Extends ethical best practices, strengthens the EU's international credibility and humanitarian leadership
- Prevents legal grey zones that might otherwise lead to reputational or operational harm

Disadvantages

- Difficult to monitor or enforce outside EU jurisdiction
- May require additional institutional resources to implement and audit such standards.

Option 8. Strengthen human oversight, transparency, and bias mitigation in AI tools

Why: Human-in-the-loop and transparency requirements under the AI Act are especially critical during crises, where decisions can be high-stakes, to ensure trust, fairness, and accountability. Bias in training data may result in unequal resource allocation or missed vulnerabilities.

How: Mandate pre-authorisation or certification of AI tools used in civil protection and emergency response (at ERCC or national level), including independent audits for fairness, explainability, and robustness; establish hybrid human-AI workflows and ethical safeguards

Advantages

- Promotes trust, contestability, and informed decision-making
- Reduces risk of harm to marginalised or underrepresented groups.

Disadvantages

- May slow down automation gains or require additional staffing and training
- Can be difficult to implement in real-time, high-pressure situations
- Requires trained personnel and ongoing coordination.

Option 9. Operationalise ethical AI through strategic data governance and coordination

Why: The effectiveness and acceptability of AI in crisis management depends on access to quality data and coordination between actors. The EU's Data Act and AI Act provide a foundation, but strategic-level governance is still fragmented.

How: Establish a dedicated EU Crisis AI Coordination Mechanism (e.g. under the ERCC) to oversee data governance, model validation, risk audits, and partner compliance during AI deployment in disasters.

Advantages

- Enables more effective and responsible use of private-sector and cross-border data
- Improves preparedness through interoperable and high-quality data inputs.

Disadvantages

- Requires multi-stakeholder cooperation, which may be hindered by competing interests or lack of trust
- Legal and ethical responsibilities can become diluted without a central coordinating authority.

References

Abels, J. (2024). Private infrastructure in geopolitical conflicts: The case of Starlink and the war in Ukraine. *European Journal of International Relations*, *30*(4), 842–866. https://doi.org/10.1177/13540661241260653

Abraham, S., Carmichael, Z., Banerjee, S., VidalMata, R., Agrawal, A., Al Islam, M. N.,…Cleland-Huang, J. (2021). Adaptive autonomy in human-on-the-loop vision-based robotics systems. *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)*, 113–120. https://doi.org/10.1109/WAIN52551.2021.00025

Acharya, D. B., Kuppan, K., & Divya, B. (2025). Agentic AI: Autonomous intelligence for complex goals: A comprehensive survey. *IEEE Access*. https://doi.org/10.1109/ACCESS.2025.3532853

Ahmed, N., Michelin, R. A., Xue, W., Ruj, S., Malaney, R., Kanhere, S. S.,…Jha, S. K. (2020). A survey of COVID-19 contact tracing apps. *IEEE Access*, *8*, 134577–134601. https://doi.org/10.1109/ACCESS.2020.3010226

Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G.,…Hoos, H. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, *53*(8), 18–28. https://doi.org/10.1109/MC.2020.2996587

Alsamhi, S. H., Kumar, S., Hawbani, A., Shvetsov, A. V., Zhao, L., & Guizani, M. (2024). Synergy of human-centered AI and cyber-physical-social systems for enhanced cognitive situation awareness: Applications, challenges and opportunities. *Cognitive Computation*, *16*(5), 2735–2755. https://doi.org/10.1007/s12559-024-10271-7

Altay, N., & Labonte, M. (2014). Challenges in humanitarian information management and exchange: Evidence from Haiti. *Disasters*, *38*(s1), S50–S72. https://doi.org/10.1111/disa.12052

Amoroso, D., & Tamburrini, G. (2020). Autonomous weapons systems and meaningful human control: Ethical and legal issues. *Current Robotics Reports*, *1*(4), 187–194. https://doi.org/10.1007/s43154-020-00024-3

An, J., Huang, D., Lin, C., & Tai, M. (2024). Measuring gender and racial biases in large language models. *arXiv Preprint arXiv:2403.15281*. https://doi.org/10.1093/pnasnexus/pgaf089

Anom, B. Y. (2022). The ethical dilemma of mobile phone data monitoring during COVID-19: The case for South Korea and the United States. *Journal of Public Health Research*, *11*(3), 22799036221102491. https://doi.org/10.1177/22799036221102491

Asnaning, A. R., & Putra, S. D. (2018). Flood early warning system using cognitive artificial intelligence: The design of AWLR sensor. *2018 International Conference on Information Technology Systems and Innovation (ICITSI)*, 165–170. https://doi.org/10.1109/ICITSI.2018.8695948

Aylett-Bullock, J., Gilman, R. T., Hall, I., Kennedy, D., Evers, E. S., Katta, A.,…Ariqi, L. (2022). Epidemiological modelling in refugee and internally displaced people settlements: Challenges and ways forward. *BMJ Global Health*, *7*(3), e007822. https://doi.org/10.1136/bmjgh-2021-007822

Ayoub, J., Yang, X. J., & Zhou, F. (2021). Combat COVID-19 infodemic using explainable natural language processing models. *Information Processing & Management*, *58*(4), 102569. https://doi.org/10.1016/j.ipm.2021.102569

Bach, T. A., Khan, A., Hallock, H., Beltrão, G., & Sousa, S. (2024). A systematic literature review of user trust in AI-enabled systems: An HCI perspective. *International Journal of Human–Computer Interaction*, *40*(5), 1251–1266. https://doi.org/10.1080/10447318.2022.2138826

Bak-Coleman, J. B., Kennedy, I., Wack, M., Beers, A., Schafer, J. S., Spiro, E. S.,…West, J. D. (2022). Combining interventions to reduce the spread of viral misinformation. *Nature Human Behaviour*, *6*(10), 1372–1380. https://doi.org/10.1038/s41562-022-01388-6

Banerjee, S., Agarwal, A., & Singla, S. (2025). LLMs will always hallucinate, and we need to live with this. *Intelligent Systems Conference*, 624–648. https://doi.org/10.1007/978-3-031-99965-9_39

Bansal, G., Nushi, B., Kamar, E., Horvitz, E., & Weld, D. S. (2021). Is the most accurate AI the best teammate? Optimizing AI for teamwork. *Proceedings of the AAAI Conference on Artificial Intelligence*, *35*(13), 11405–11414. https://doi.org/10.1609/aaai.v35i13.17359

Barnawi, A., Chhikara, P., Tekchandani, R., Kumar, N., & Alzahrani, B. (2021). Artificial intelligence-enabled Internet of Things-based system for COVID-19 screening using aerial thermal imaging. *Future Generation Computer Systems*, *124*, 119–132. https://doi.org/10.1016/j.future.2021.05.019

Bashir, N., Donti, P., Cuff, J., Sroka, S., Ilic, M., Sze, V.,… Olivetti, E. (2024). The climate and sustainability implications of generative AI. *An MIT Exploration of Generative AI*, *3*(7). Retrieved from https://mit-genai.pubpub.org/pub/8ulgrckc

Batool, A., Zowghi, D., & Bano, M. (2025). AI governance: A systematic literature review. *AI and Ethics*, 1–15. https://doi.org/10.1007/s43681-024-00653-w

Bäumler, J., Blöcher, L., Frey, L.-J., Chen, X., Bayer, M., & Reuter, C. (2025). A survey of machine learning models and datasets for the multi-label classification of textual hate speech in English. *arXiv Preprint arXiv:2504.08609*. https://doi.org/10.48550/arXiv.2504.08609

Bayer, M., Neiczer, M., Samsinger, M., Buchhold, B., & Reuter, C. (2024). XAI-attack: Utilizing explainable AI to find incorrectly learned patterns for black-box adversarial example creation. *Proceedings of the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024)*, 17725–17738. Retrieved from https://aclanthology.org/2024.lrec-main.1542/

Benaben, F., Fertier, A., Montarnal, A., Mu, W., Jiang, Z., Truptil, S.,…Wang, T. (2020). An AI framework and a metamodel for collaborative situations: Application to crisis management contexts. *Journal of Contingencies and Crisis Management*, *28*(3), 291–306. https://doi.org/10.1111/1468-5973.12310

Betke, H., Peitzsch, S., Boldt, J., Reimann, D., & Kox, T. (2024). Chatbot based public sensoring to improve situational awareness. *Proceedings of the International ISCRAM Conference*. https://doi.org/10.59297/h3r7bp59

Bhadauria, P. K. S. (2024). Comprehensive review of AI and ML tools for earthquake damage assessment and retrofitting strategies. *Earth Science Informatics*, *17*(5), 3945–3962. https://doi.org/10.1007/s12145-024-01431-2

Bharosa, N., Lee, J., & Janssen, M. (2010). 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

Bhatia, M., Ahanger, T. A., & Manocha, A. (2023). Artificial intelligence based real-time earthquake prediction. *Engineering Applications of Artificial Intelligence*, *120*, 105856. https://doi.org/10.1016/j.engappai.2023.105856

Bhatnagar, T., Omar, M., Orlic, D., Smith, J., Holloway, C., & Kett, M. (2025). Bridging AI and humanitarianism: An HCI-informed framework for responsible AI adoption. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3706598.3713184

Bhuiyan, M. M., Whitley, H., Horning, M., Lee, S. W., & Mitra, T. (2021). Designing transparency cues in online news platforms to promote trust: Journalists' & consumers' perspectives. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–31. https://doi.org/10.1145/3479539

Bilquise, G., Ibrahim, S., & Shaalan, K. (2022). Emotionally intelligent chatbots: A systematic literature review. *Human Behavior and Emerging Technologies*, *2022*(1), 9601630. https://doi.org/10.1155/2022/9601630

Blakey, E. (2024). The day data transparency died: How Twitter/X cut off access for social research. *Contexts*, *23*(2), 30–35. https://doi.org/10.1177/15365042241252125

Bleick, M., Feldhus, N., Burchardt, A., & Möller, S. (2024). German voter personas can radicalize LLM chatbots via the echo chamber effect. *Proceedings of the 17th International Natural Language Generation Conference*, 153–164. https://doi.org/10.18653/v1/2024.inlg-main.13

Boin, A., Ekengren, M., & Rhinard, M. (2016). The study of crisis management. In *Routledge handbook of security studies* (pp. 447–456). Routledge. Retrieved from https://www.researchgate.net/publication/48328170_The_Study_of_Crisis_Management

Bolón-Canedo, V., Morán-Fernández, L., Cancela, B., & Alonso-Betanzos, A. (2024). A review of green artificial intelligence: Towards a more sustainable future. *Neurocomputing*, *599*, 128096. https://doi.org/10.1016/j.neucom.2024.128096

Borges, M. R., Canós, J. H., Penadés, M. C., Labaka, L., Bañuls, V. A., & Hernantes, J. (2023). Toward a taxonomy for classifying crisis information management systems. In *Disaster management and information technology: Professional response and recovery management in the age of disasters* (pp. 409–433). Springer. https://doi.org/10.1007/978-3-031-20939-0_19

Bradshaw, J. M., Dignum, V., Jonker, C., & Sierhuis, M. (2012). Human-agent-robot teamwork. *IEEE Intelligent Systems*, *27*(2), 8–13. https://doi.org/10.1109/MIS.2012.37

Brewczyńska, M. (2020). Poland: Policing quarantine via app. In *Data justice and COVID-19: Global perspectives* (pp. 232–239). Meatspace Press. https://research.tilburguniversity.edu/en/publications/poland-policing-quarantine-via-app/

Buçinca, Z. (2024). Optimizing decision-makers' intrinsic motivation for effective human-AI decision-making. *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, 1–5. https://dl.acm.org/doi/10.1145/3613905.3638179

Buçinca, Z., Swaroop, S., Paluch, A. E., Doshi-Velez, F., & Gajos, K. Z. (2025). Contrastive explanations that anticipate human misconceptions can improve human decision-making skills. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–25. https://doi.org/10.1145/3706598.3713229

Burnap, P., & Williams, M. L. (2016). Us and them: Identifying cyber hate on Twitter across multiple protected characteristics. *EPJ Data Science*, *5*(1), 11. https://doi.org/10.1140/epjds/s13688-016-0072-6

Caldarelli, G., Arcaute, E., Barthelemy, M., Batty, M., Gershenson, C., Helbing, D.,…Rozenblat, C. (2023). The role of complexity for digital twins of cities. *Nature Computational Science*, *3*(5), 374–381. https://doi.org/10.1038/s43588-023-00431-4

Calvert, S., Johnsen, S., & George, A. (2024). Designing automated vehicle and traffic systems towards meaningful human control. In *Research handbook on meaningful human control of artificial intelligence systems* (pp. 162–187). Edward Elgar Publishing. https://doi.org/10.4337/9781802204131.00018

Canós, J. H., Alonso, G., & Jaén, J. (2004). A multimedia approach to the efficient implementation and use of emergency plans. *IEEE Multimedia*, *11*(3), 106–110. https://doi.org/10.1109/MMUL.2004.2

References

Caramancion, K. M. (2023). Harnessing the power of ChatGPT to decimate mis/disinformation: Using ChatGPT for fake news detection. *2023 IEEE World AI IoT Congress (AIIoT)*, 0042–0046. https://doi.org/10.1109/AIIoT58121.2023.10174450

Casali, Y., Aydin, N. Y., & Comes, T. (2022). Machine learning for spatial analyses in urban areas: A scoping review. *Sustainable Cities and Society*, *85*, 104050. https://doi.org/10.1016/j.scs.2022.104050

Cavalcante Siebert, L., Lupetti, M. L., Aizenberg, E., Beckers, N., Zgonnikov, A., Veluwenkamp, H.,…Jonker, C. M. (2023). Meaningful human control: Actionable properties for AI system development. *AI and Ethics*, *3*(1), 241–255. https://doi.org/10.1007/s43681-022-00167-3

Chamberlain, H. R., Darin, E., Adewole, W. A., Jochem, W. C., Lazar, A. N., & Tatem, A. J. (2024). Building footprint data for countries in Africa: To what extent are existing data products comparable? *Computers, Environment and Urban Systems*, *110*, 102104. https://doi.org/10.1016/j.compenvurbsys.2024.102104

Charlton-Perez, A. J., Dacre, H. F., Driscoll, S., Gray, S. L., Harvey, B., Harvey, N. J.,…Vandaele, R. (2024). Do AI models produce better weather forecasts than physics-based models? A quantitative evaluation case study of Storm Ciarán. *Npj Climate and Atmospheric Science*, *7*(1), 93. https://doi.org/10.1038/s41612-024-00638-w

Chaudhari, S., Aggarwal, P., Murahari, V., Rajpurohit, T., Kalyan, A., Narasimhan, K.,…Silva, B. (2025). RLHF deciphered: A critical analysis of reinforcement learning from human feedback for LLMs. *ACM Computing Surveys*, *58*(2), 1–37. https://doi.org/10.1145/3743127

Chaudhuri, N., & Bose, I. (2020). Exploring the role of deep neural networks for post-disaster decision support. *Decision Support Systems*, *130*, 113234. https://doi.org/10.1016/j.dss.2019.113234

Chen, C., & Shu, K. (2023). Can LLM-generated misinformation be detected? *arXiv Preprint arXiv:2309.13788*. https://doi.org/10.48550/arXiv.2309.13788

Chen, C., & Shu, K. (2024). Combating misinformation in the age of LLMs: Opportunities and challenges. *AI Magazine*, *45*(3), 354–368. https://doi.org/10.1002/aaai.12188

Chhabra, A., & Vishwakarma, D. K. (2023). A literature survey on multimodal and multilingual automatic hate speech identification. *Multimedia Systems*, *29*(3), 1203–1230. https://doi.org/10.1007/s00530-023-01051-8

Chiu, K.-L., Collins, A., & Alexander, R. (2021). Detecting hate speech with gpt-3. *arXiv Preprint arXiv:2103.12407*. https://doi.org/10.1075/ps.21010.chu

Choung, H., David, P., & Ross, A. (2023). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human–Computer Interaction*, *39*(9), 1727–1739. https://doi.org/10.1080/10447318.2022.2050543

Chun, K. P., Octavianti, T., Dogulu, N., Tyralis, H., Papacharalampous, G., Rowberry, R.,… Migliari, W. (2025). Transforming disaster risk reduction with AI and big data: Legal and interdisciplinary perspectives. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *15*(2), e70011. https://doi.org/10.1002/widm.70011

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

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. https://doi.org/10.1111/poms.13236

Comfort, L. K. (2007). Crisis management in hindsight: Cognition, communication, coordination, and control. *Public Administration Review*, *67*, 189–197. https://doi.org/10.1111/j.1540-6210.2007.00827.x

Conges, A., Evain, A., Benaben, F., Chabiron, O., & Rebiere, S. (2020). Crisis management exercises in virtual reality. *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, 87–92. https://doi.org/10.1109/VRW50115.2020.00022

Coughlan de Perez, E., van den Hurk, B., Van Aalst, M. K., Jongman, B., Klose, T., & Suarez, P. (2015). Forecast-based financing: An approach for catalyzing humanitarian action based on extreme weather and climate forecasts. *Natural Hazards and Earth System Sciences*, *15*(4), 895–904. https://doi.org/10.5194/nhess-15-895-2015

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*(4), 491–502. https://doi.org/10.1007/s10708-014-9597-z

Crowston, K., & Bolici, F. (2025). Deskilling and upskilling with generative AI systems. *Information Research an International Electronic Journal, 30(iConf), 1009–1023*. https://doi.org/10.47989/ir30iConf47143

Cummings, M. M. (2014). Man versus machine or man + machine? *IEEE Intelligent Systems*, *29*(5), 62–69. https://doi.org/10.1109/MIS.2014.87

Data centres & networks. (n.d.). IEA. Retrieved 2 September 2025, from https://www.iea.org/energy-system/buildings/data-centres-and-data-transmission-networks

Data sharing in an urgent situation or in an emergency. (2024, November 19). ICO. Retrieved from https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/data-sharing/data-sharing-a-code-of-practice/data-sharing-in-an-urgent-situation-or-in-an-emergency/

Davidson, T., Bhattacharya, D., & Weber, I. (2019). Racial bias in hate speech and abusive language detection datasets. *arXiv Preprint arXiv:1905.12516*. https://doi.org/10.18653/v1/W19-3504

De Clercq, D., Xu, L., de Ruiter, M., van den Homberg, M., van der Velde, M., Hall, J.,…Mahdi, A. (2025). Towards optimal anticipatory action: Maximizing the effectiveness of agricultural early warning systems with operations research. *International Journal of Disaster Risk Reduction*, 105249. https://doi.org/10.1016/j.ijdrr.2025.105249

Deleforge, A., Di Carlo, D., Strauss, M., Serizel, R., & Marcenaro, L. (2019). Audio-based search and rescue with a drone: Highlights from the IEEE signal processing cup 2019 student competition [SP competitions]. *IEEE Signal Processing Magazine*, *36*(5), 138–144. https://doi.org/10.1109/MSP.2019.2924687

Deléglise, H., Interdonato, R., Bégué, A., d'Hôtel, E. M., Teisseire, M., & Roche, M. (2022). Food security prediction from heterogeneous data combining machine and deep learning methods. *Expert Systems with Applications*, *190*, 116189. https://doi.org/10.1016/j.eswa.2021.116189

Delgado, J., de Manuel, A., Parra, I., Moyano, C., Rueda, J., Guersenzvaig, A.,…Puyol, A. (2022). Bias in algorithms of AI systems developed for COVID-19: A scoping review. *Journal of Bioethical Inquiry*, *19*(3), 407–419. https://doi.org/10.1007/s11673-022-10200-z

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, *144*(1), 114. https://doi.org/10.1037/xge0000033

Ding, X., Shang, B., Xie, C., Xin, J., & Yu, F. (2025). Artificial intelligence in the COVID-19 pandemic: Balancing benefits and ethical challenges in China's response. *Humanities and Social Sciences Communications*, *12*(1), 1–19. https://doi.org/10.1057/s41599-025-04564-x

Directorate-General for Parliamentary Research Services (European Parliament), & Boucher, P. (2020). *Artificial intelligence: How does it work, why does it matter, and what we can do about it?* Publications Office of the European Union. https://data.europa.eu/doi/10.2861/44572

Drones in humanitarian action: A guide to the use of airborne systems in humanitarian crises. *ReliefWeb*. (2016, December 2). Retrieved from https://reliefweb.int/report/world/drones-humanitarian-action-guide-use-airborne-systems-humanitarian-crises

Duan, Y., Edwards, J. S., & Dwivedi, Y. K. (2019). Artificial intelligence for decision making in the era of Big Data–evolution, challenges and research agenda. *International Journal of Information Management*, *48*, 63–71. https://**doi**.org/10.1016/j.ijinfomgt.2019.01.021

Dwivedi, R., Dave, D., Naik, H., Singhal, S., Omer, R., Patel, P.,…Morgan, G. (2023). Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Computing Surveys*, *55*(9), 1–33. https://doi.org/10.1145/3561048

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K.,…Ahuja, M. (2023). Opinion paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Ebener, S., Castro, F., & Dimailig, L. A. (2014). *Increasing availability, quality, and accessibility of common and fundamental operational datasets to support disaster risk reduction and emergency management in the Philippines*. Manila.

Edwards, T. E., & Lee, P. U. (2017). Towards designing graceful degradation into trajectory-based operations: A human-machine system integration approach. *17th AIAA Aviation Technology, Integration, and Operations Conference*, 4487. https://doi.org/10.2514/6.2017-4487

Eini, M., Kaboli, H. S., Rashidian, M., & Hedayat, H. (2020). Hazard and vulnerability in urban flood risk mapping: Machine learning techniques and considering the role of urban districts. *International Journal of Disaster Risk Reduction*, *50*, 101687. https://doi.org/10.1016/j.ijdrr.2020.101687

Eisenstein, J. (2019). *Introduction to natural language processing*. MIT Press.

Ekelhof, M. (2019). Moving beyond semantics on autonomous weapons: Meaningful human control in operation. *Global Policy*, *10*(3), 343–348. https://doi.org/10.1111/1758-5899.12665

El-Masri, A., Riedl, M. J., & Woolley, S. (2022). Audio misinformation on WhatsApp: A case study from Lebanon. *Harvard Kennedy School Misinformation Review*, *3*(4), 1–13. https://doi.org/10.37016/mr-2020-102

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors*, *59*(1), 5–27. https://doi.org/10.1177/0018720816681350

Esparza, M., Li, B., Ma, J., & Mostafavi, A. (2025). AI meets natural hazard risk: A nationwide vulnerability assessment of data centers to natural hazards and power outages. *International Journal of Disaster Risk Reduction*, 105583. https://doi.org/10.1016/j.ijdrr.2025.105583

Espeholt, L., Agrawal, S., Sønderby, C., Kumar, M., Heek, J., Bromberg, C.,…Hickey, J. (2022). Deep learning for twelve-hour precipitation forecasts. *Nature Communications*, *13*(1), 5145. https://doi.org/10.1038/s41467-022-32483-x

European Group on Ethics. (2022). Values in times of crisis – Strategic crisis management in the EU. *Publications Office of the European Union*. https://data.europa.eu/doi/10.2777/79910

Fan, C., Zhang, C., Yahja, A., & Mostafavi, A. (2021). Disaster City Digital Twin: A vision for integrating artificial and human intelligence for disaster management. *International Journal of Information Management*, *56*, 102049. https://doi.org/10.1016/j.ijinfomgt.2019.102049

Fan, Y., Wang, Z., Deng, S., Lv, H., & Wang, F. (2022). The function and quality of individual epidemic prevention and control apps during the COVID-19 pandemic: A systematic review of Chinese apps. *International Journal of Medical Informatics*, *160*, 104694. https://doi.org/10.1016/j.ijmedinf.2022.104694

Fathi, R., & Fiedrich, F. (2022). Social media analytics by virtual operations support teams in disaster management: Situational awareness and actionable information for decision-makers. *Frontiers in Earth Science*, *10*, 941803. https://doi.org/10.3389/feart.2022.941803

Finn, M., & Oreglia, E. (2016). A fundamentally confused document: Situation reports and the work of producing humanitarian information. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, 1349–1362. https://doi.org/10.1145/2818048.2820031

Fisher, C. W., & Kingma, B. R. (2001). Criticality of data quality as exemplified in two disasters. *Information & Management*, *39*(2), 109–116. https://doi.org/10.1016/S0378-7206(01)00083-0

Fitts, P. M. (1951). Human engineering for an effective air-navigation and traffic-control system. *National Research Council, Div. Of.* https://psycnet.apa.org/record/1952-01751-000

Freed, D., Bazarova, N. N., Consolvo, S., Han, E. J., Kelley, P. G., Thomas, K., & Cosley, D. (2023). Understanding digital-safety experiences of youth in the US. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3544548.3581128

Fu, D., Ban, Y., Tong, H., Maciejewski, R., & He, J. (2022). Disco: Comprehensive and explainable disinformation detection. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 4848–4852. https://doi.org/10.1145/3511808.3557202

Gevaert, C. M., Buunk, T., & Van Den Homberg, M. J. (2024). Auditing geospatial datasets for biases: Using global building datasets for disaster risk management. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *17*, 12579–12590. https://doi.org/10.1109/JSTARS.2024.3422503

Gevaert, C. M., Carman, M., Rosman, B., Georgiadou, Y., & Soden, R. (2021). Fairness and accountability of AI in disaster risk management: Opportunities and challenges. *Patterns*, *2*(11). https://doi.org/10.1016/j.patter.2021.100363

Ghaffarian, S., & Kerle, N. (2019). Towards post-disaster debris identification for precise damage and recovery assessments from UAV and satellite images. *4th ISPRS Geospatial Week 2019*, 297–302. https://doi.org/10.5194/isprs-archives-XLII-2-W13-297-2019

Goyal, N., Kivlichan, I. D., Rosen, R., & Vasserman, L. (2022). Is your toxicity my toxicity? Exploring the impact of rater identity on toxicity annotation. *Proceedings of the ACM on Human-Computer Interaction*, *6*(CSCW2), 1–28. https://doi.org/10.1145/3555088

Grandhi, S., Plotnick, L., & Hiltz, S. R. (2021). By the crowd and for the crowd: Perceived utility and willingness to contribute to trustworthiness indicators on social media. *Proceedings of the ACM on Human-Computer Interaction*, *5*(GROUP), 1–24. https://doi.org/10.1145/3463930

Grass, E., Ortmann, J., Balcik, B., & Rei, W. (2023). A machine learning approach to deal with ambiguity in the humanitarian decision-making. *Production and Operations Management*, *32*(9), 2956–2974. https://doi.org/10.1111/poms.14018

Green AI – Communications of the ACM. (2020, December 1). https://cacm.acm.org/research/green-ai/

Gstrein, O. J., Haleem, N., & Zwitter, A. (2024). General-purpose AI regulation and the European Union AI Act. *Internet Policy Review*, *13*(3), 1–26. https://doi.org/10.14763/2024.3.1790

Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., & Yang, G.-Z. (2019). XAI: Explainable artificial intelligence. *Science Robotics*, *4*(37), eaay7120. https://doi.org/10.1126/scirobotics.aay7120

Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *International Conference on Machine Learning*, 1321–1330. Retrieved from https://proceedings.mlr.press/v70/guo17a.html

Gupta, R., & Shah, M. (2021). Rescuenet: Joint building segmentation and damage assessment from satellite imagery. *2020 25th International Conference on Pattern Recognition (ICPR)*, 4405–4411. https://doi.org/10.1109/ICPR48806.2021.9412295

Hartmann, D., Oueslati, A., Staufer, D., Pohlmann, L., Munzert, S., & Heuer, H. (2025). Lost in moderation: How commercial content moderation APIs over-and under-moderate group-targeted hate speech and linguistic variations. *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–26. https://doi.org/10.1145/3706598.3713998

Hartwig, K., Sandler, R., & Reuter, C. (2024). Navigating misinformation in voice messages: Identification of user-centered features for digital interventions. *Risk, Hazards & Crisis in Public Policy*, *15*(2), 203–235. https://doi.org/10.1002/rhc3.12296

Hayes, P., & Kelly, S. (2018). Distributed morality, privacy, and social media in natural disaster response. *Technology in Society*, *54*, 155–167. https://doi.org/10.1016/j.techsoc.2018.05.003

Haykal, D., Goldust, M., Cartier, H., & Treacy, P. (2025). AI in humanitarian healthcare: A game changer for crisis response. *Frontiers in Artificial Intelligence*, *8*, 1627773. https://doi.org/10.3389/frai.2025.1627773

He, J., Rungta, M., Koleczek, D., Sekhon, A., Wang, F. X., & Hasan, S. (2024). Does prompt formatting have any impact on LLM performance? *arXiv Preprint arXiv:2411.10541*. https://doi.org/10.48550/arXiv.2411.10541

Hemmer, P., Schemmer, M., Kühl, N., Vössing, M., & Satzger, G. (2025). Complementarity in human-AI collaboration: Concept, sources, and evidence. *European Journal of Information Systems*, 1–24. https://doi.org/10.1080/0960085X.2025.2475962

Hernandez, K., & Roberts, T. (2020). Predictive analytics in humanitarian action: A preliminary mapping and analysis. *K4D Emerging Issues Report 33*. Brighton, UK: Institute of Development Studies. Retrieved from: https://www.gov.uk/research-for-development-outputs/predictive-analytics-in-humanitarian-action-a-preliminary-mapping-and-analysis

Herrmann, T., & Pfeiffer, S. (2023). Keeping the organization in the loop: A socio-technical extension of human-centered artificial intelligence. *Ai & Society*, *38*(4), 1523–1542. https://doi.org/10.1007/s00146-022-01391-5

Hida, R., Kaneko, M., & Okazaki, N. (2024). Social bias evaluation for large language models requires prompt variations. *arXiv Preprint arXiv:2407.03129*. https://doi.org/10.18653/v1/2025.findings-emnlp.783

Hildebrandt, J. R., Ziefle, M., & Calero Valdez, A. (2022). Entscheidungsautonomie und KI-Methodische Hinweise zur Untersuchung von KI-Nutzung in Sicherheitsbehörden. *Mensch Und Computer 2022-Workshopband*, 10.18420/muc2022-mci-ws10-230. https://doi.org/10.18420/muc2022-mci-ws10-230

Hille, E. M., Hummel, P., & Braun, M. (2023). Meaningful human control over AI for health? A review. *Journal of Medical Ethics*. https://doi.org/10.1136/jme-2023-109095

Hoang, M.-T. O., Grøntved, K. A. R., van Berkel, N., Skov, M. B., Christensen, A. L., & Merritt, T. (2023). Drone swarms to support search and rescue operations: Opportunities and challenges. *Cultural Robotics: Social Robots and Their Emergent Cultural Ecologies*, 163–176. https://doi.org/10.1007/978-3-031-28138-9_11

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Hoffmann, J., Bauer, P., Sandu, I., Wedi, N., Geenen, T., & Thiemert, D. (2023). *Destination Earth–A digital twin in support of climate services*. Elsevier. https://doi.org/10.1016/j.cliser.2023.100394

How can AI strengthen disaster preparedness in Europe? | UCP Knowledge Network. (n.d.). Retrieved 2 September 2025, from https://civil-protection-knowledge-network.europa.eu/news/how-can-ai-strengthen-disaster-preparedness-europe

Huang, L., Yu, W., Ma, W., Zhong, W., Feng, Z., Wang, H., Chen, Q., Peng, W., Feng, X., & Qin, B. (2025). A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, *43*(2), 1–55. https://doi.org/10.1145/3729422

Islam, M. A., Rashid, S. I., Hossain, N. U. I., Fleming, R., & Sokolov, A. (2023). An integrated convolutional neural network and sorting algorithm for image classification for efficient flood disaster management. *Decision Analytics Journal*, *7*, 100225. https://doi.org/10.1016/j.dajour.2023.100225

Jacob, C., Kerrigan, P., & Bastos, M. (2025). The chat-chamber effect: Trusting the AI hallucination. *Big Data & Society*, *12*(1), 20539517241306345. https://doi.org/10.1177/20539517241306345

Jahan, M. S., & Oussalah, M. (2023). A systematic review of hate speech automatic detection using natural language processing. *Neurocomputing*, *546*, 126232. https://doi.org/10.1016/j.neucom.2023.126232

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. https://doi.org/10.1145/2629559

Jiang, B., Tan, Z., Nirmal, A., & Liu, H. (2024). Disinformation detection: An evolving challenge in the age of LLMs. *Proceedings of the 2024 Siam International Conference on Data Mining (SDM)*, 427–435. https://doi.org/10.1137/1.9781611978032.50

Johnson, M., Albizri, A., Harfouche, A., & Tutun, S. (2023). Digital transformation to mitigate emergency situations: Increasing opioid overdose survival rates through explainable artificial intelligence. *Industrial Management & Data Systems*, *123*(1), 324–344. https://doi.org/10.1108/IMDS-04-2021-0248

Jones, A., Kuehnert, J., Fraccaro, P., Meuriot, O., Ishikawa, T., Edwards, B.,…Assefa, S. (2023). AI for climate impacts: Applications in flood risk. *Npj Climate and Atmospheric Science*, *6*(1), 63. https://doi.org/10.1038/s41612-023-00388-1

Jones, D., Fischer, J. E., Rodden, T., Reece, S., Ramchurn, S. D., & Allen, S. (2015). Augmenting the bird table: Developing technological support for disaster response. *Procedia Engineering*, *107*, 54–58. https://doi.org/10.1016/j.proeng.2015.06.058

Jones, L., A. Constas, M., Matthews, N., & Verkaart, S. (2021). Advancing resilience measurement. *Nature Sustainability*, *4*(4), 288–289. https://doi.org/10.1038/s41893-020-00642-x

Kaelbling, L. P., Littman, M. L., & Moore, A. W. (1996). Reinforcement learning: A survey. *Journal of Artificial Intelligence Research*, *4*, 237–285. https://doi.org/10.1613/jair.301

Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31. https://doi.org/10.1287/mnsc.39.1.17

Kaliyar, R. K., Goswami, A., & Narang, P. (2019). Multiclass fake news detection using ensemble machine learning. *2019 IEEE 9th International Conference on Advanced Computing (IACC)*, 103–107. https://doi.org/10.1109/IACC48062.2019.8971579

Karam, S., Nex, F., Chidura, B. T., & Kerle, N. (2022). Microdrone-based indoor mapping with graph slam. *Drones*, *6*(11), 352. https://doi.org/10.3390/drones6110352

Kaufhold, M.-A., Rupp, N., Reuter, C., & Habdank, M. (2020). Mitigating information overload in social media during conflicts and crises: Design and evaluation of a cross-platform alerting system. *Behaviour & Information Technology*, *39*(3), 319–342. https://doi.org/10.1080/0144929X.2019.1620334

Kazemi, A. (n.d.). *Futurium | European AI Alliance - CrisisAI: A novel hybrid AI system for crisis management*. Retrieved 2 September 2025, from https://futurium.ec.europa.eu/de/european-ai-alliance/document/crisisai-novel-hybrid-ai-system-crisis-management

Kejriwal, M., & Zhou, P. (2020). On detecting urgency in short crisis messages using minimal supervision and transfer learning. *Social Network Analysis and Mining*, *10*(1), 58. https://doi.org/10.1007/s13278-020-00670-7

Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, *30*. https://doi.org/10.48550/arXiv.1703.04977

Kikon, A., & Deka, P. C. (2022). Artificial intelligence application in drought assessment, monitoring and forecasting: A review. *Stochastic Environmental Research and Risk Assessment*, *36*(5), 1197–1214. https://doi.org/10.1007/s00477-021-02129-3

Kim, K., & Boulanin, V. (2023). *Artificial intelligence for climate security: Possibilities and challenges*. https://doi.org/10.55163/QDSE8934

Kirchner, J., & Reuter, C. (2020). Countering fake news: A comparison of possible solutions regarding user acceptance and effectiveness. *Proceedings of the ACM on Human-Computer Interaction*, *4*(CSCW2), 1–27. https://doi.org/10.1145/3415211

Kiritchenko, S., Nejadgholi, I., & Fraser, K. C. (2021). Confronting abusive language online: A survey from the ethical and human rights perspective. *Journal of Artificial Intelligence Research*, *71*, 431–478. https://doi.org/10.1613/jair.1.12590

Kjærum, A., & Madsen, B. S. (2025). Pushing the boundaries of anticipatory action using machine learning. *Data & Policy*, *7*, e8. https://doi.org/10.1017/dap.2024.88

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

Kox, T., Harrison, S., Ziegler, F., & Gerhold, L. (2025). Perceptions, hopes, and concerns regarding the possibilities of artificial intelligence in weather warning contexts. *International Journal of Disaster Risk Reduction*, 105817. https://doi.org/10.1016/j.ijdrr.2025.105817

Kudina, O. (2021). Bridging privacy and solidarity in COVID-19 contact-tracing apps through the sociotechnical systems perspective. *Glimpse*, *22*(2), 43–54. https://doi.org/10.5840/glimpse202122224

Kuglitsch, M. M., Pelivan, I., Ceola, S., Menon, M., & Xoplaki, E. (2022). Facilitating adoption of AI in natural disaster management through collaboration. *Nature Communications*, *13*(1), 1579. https://doi.org/10.1038/s41467-022-29285-6

Kuner, C., & Marelli, M. (2017). *Handbook on data protection in humanitarian action.(International Committee of the Red Cross)*. Retrieved from https://www.icrc.org/en/data-protection-humanitarian-action-handbook

Laaksonen, S.-M. (2023). The datafication of hate speech. *86272*, *12*, 301–317. https://doi.org/10.48541/dcr.v12.18

Lai, V., Carton, S., Bhatnagar, R., Liao, Q. V., Zhang, Y., & Tan, C. (2022). Human-AI collaboration via conditional delegation: A case study of content moderation. *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, 1–18. https://doi.org/10.1145/3491102.3501999

Lam, R., Sanchez-Gonzalez, A., Willson, M., Wirnsberger, P., Fortunato, M., Alet, F., Ravuri, S., …Hu, W. (2023). Learning skillful medium-range global weather forecasting. *Science*, *382*(6677), 1416–1421. https://doi.org/10.1126/science.adi2336

Lannelongue, L., Grealey, J., & Inouye, M. (n.d.). Green algorithms: Quantifying the carbon footprint of computation. *Advanced Science*, *8, 12 (2021): 2100707.* https://doi.org/10.1002/advs.202100707

Lashkul, V. (2021). The mobile app "Dii vdoma": Advantages and challenges. *Contemporary Management*, 280-282. (Translated from Ukrainan: Лашкул, В. (2021). Мобільний застосунок «Дій вдома»: переваги та недоліки.

Lee, C.-C., Comes, T., Finn, M., & Mostafavi, A. (2022). Roadmap towards responsible AI in crisis resilience management. *arXiv Preprint arXiv:2207.09648*. https://doi.org/10.48550/arXiv.2207.09648

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, S. G., & Kim, E. (2022). Self-quarantine system and personal information privacy in South Korea. *Yonsei Medical Journal*, *63*(9), 806. https://doi.org/10.3349/ymj.2022.63.9.806

Leslie, D., Mazumder, A., Peppin, A., Wolters, M. K., & Hagerty, A. (2021). Does "AI" stand for augmenting inequality in the era of COVID-19 healthcare? *Bmj*, *372*. https://doi.org/10.2139/ssrn.3837493

Lettieri, N., Guarino, A., Zaccagnino, R., & Malandrino, D. (2023). Keeping judges in the loop: A human–machine collaboration strategy against the blind spots of AI in criminal justice. *Soft Computing*, *27*(16), 11275–11293. https://doi.org/10.1007/s00500-023-08604-z

Lorini, V., Purohit, H., Castillo, C., Gomes, J., Peterson, S., Hugues, A., Oliveira, D.,…Panizio, E. (2024). *3rd Workshop "Social media for disaster risk management: Researchers meet practitioners"*. https://doi.org/10.2760/86922

Lucas, J., Uchendu, A., Yamashita, M., Lee, J., Rohatgi, S., & Lee, D. (2023). Fighting fire with fire: The dual role of LLMs in crafting and detecting elusive disinformation. *arXiv Preprint arXiv:2310.15515*. https://doi.org/10.18653/v1/2023.emnlp-main.883

Maletzki, C., Elsenbast, C., & Reuter-Oppermann, M. (2024). Towards human-AI interaction in medical emergency call handling. *57th Hawaii International Conference on System Sciences, HICSS 2024*, 3374–3383. https://doi.org/10.24251/HICSS.2024.407

Mallapaty, S. (2025). 'Omg, did PubMed go dark?' Blackout stokes fears about database's future. *Nature*, *639*(8054), 288–288. https://doi.org/10.1038/d41586-025-00674-3

Mancia, D. (2024, November 8). Green algorithms: The environmental cost of code. *IEEE Computer Society*. Retrieved from https://www.computer.org/publications/tech-news/trends/environmental-cost-of-code/

Mandal, D., Zou, L., Wilkho, R. S., Baig, F., Abedin, J., Zhou, B., Cai, H., Gharaibeh, N., & Lam, N. (2024). Prime: A cybergis platform for resilience inference measurement and enhancement. *Computers, Environment and Urban Systems*, *114*, 102197. https://doi.org/10.1016/j.compenvurbsys.2024.102197

Manzini, T., Perali, P., Tripathi, J., & Murphy, R. R. (2025). Now you see it, now you don't: Damage label agreement in drone & satellite post-disaster imagery. *Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency*, 1998–2008. https://doi.org/10.1145/3715275.3732135

Mardani, M., Brenowitz, N., Cohen, Y., Pathak, J., Chen, C.-Y., Liu, C.-C.,…Subramaniam, A. (2025). Residual corrective diffusion modeling for km-scale atmospheric downscaling. *Communications Earth & Environment*, *6*(1), 124. https://doi.org/10.1038/s43247-025-02042-5

Marelli, M. (2024). *Handbook on data protection in humanitarian action*. https://doi.org/10.1017/9781009414630

Marzocchi, W. (2012). Putting science on trial. *Physics World*, *25*(12), 17. https://doi.org/10.1088/2058-7058/25/12/27

Mbunge, E. (2020). Integrating emerging technologies into COVID-19 contact tracing: Opportunities, challenges and pitfalls. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, *14*(6), 1631–1636. https://doi.org/10.1016/j.dsx.2020.08.029

McElhinney, H., & Spencer, S. (2024, March 11). *The clock is ticking to create minimum standards*. The New Humanitarian. Retrieved from https://www.thenewhumanitarian.org/opinion/2024/03/11/build-guardrails-humanitarian-ai

Mendonça, D., Beroggi, G. E., 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. https://doi.org/10.1016/j.ssci.2005.12.006

Mendonça, D., Jefferson, T., & Harrald, J. (2007). Collaborative adhocracies and mix-and-match technologies in emergency management. *Communications of the ACM*, *50*(3), 44–49. https://doi.org/10.1145/1226736.1226764

Meske, C., & Bunde, E. (2023). Design principles for user interfaces in AI-Based decision support systems: The case of explainable hate speech detection. *Information Systems Frontiers*, *25*(2), 743–773. https://doi.org/10.1007/s10796-021-10234-5

Middleton, S. E., Middleton, L., & Modafferi, S. (2013). Real-time crisis mapping of natural disasters using social media. *IEEE Intelligent Systems*, *29*(2), 9–17. https://doi.org/10.1109/MIS.2013.126

Migliorini, M., Hagen, J. S., Mihaljević, J., Mysiak, J., Rossi, J.-L., Siegmund, A., Meliksetian, K., & Guha Sapir, D. (2019). Data interoperability for disaster risk reduction in Europe. *Disaster Prevention and Management: An International Journal*, *28*(6), 804–816. https://doi.org/10.1108/DPM-09-2019-0291

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B.,…Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Moon, S.-H., Kim, Y.-H., Lee, Y. H., & Moon, B.-R. (2019). Application of machine learning to an early warning system for very short-term heavy rainfall. *Journal of Hydrology*, *568*, 1042–1054. https://doi.org/10.1016/j.jhydrol.2018.11.060

Mosier, K. L., & Skitka, L. J. (1999). Automation use and automation bias. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(3), 344–348. https://doi.org/10.1177/154193129904300346

Muhren, W. J., & Van de Walle, B. (2010). A call for sensemaking support systems in crisis management. In *Interactive collaborative information systems* (pp. 425–452). Springer. https://doi.org/10.1007/978-3-642-11688-9_16

Nahavandi, S. (2017). Trusted autonomy between humans and robots: Toward human-on-the-loop in robotics and autonomous systems. *IEEE Systems, Man, and Cybernetics Magazine*, *3*(1), 10–17. https://doi.org/10.1109/MSMC.2016.2623867

Narang, P. (2019). Multiclass fake news detection using ensemble machine learning. *Conference: 2019 IEEE 9th International Conference on Advanced Computing (IACC)*. https://doi.org/10.1109/IACC48062.2019.8971579

References

Nex, F., Duarte, D., Tonolo, F. G., & Kerle, N. (2019). Structural building damage detection with deep learning: Assessment of a state-of-the-art CNN in operational conditions. *Remote Sensing*, *11*(23), 2765. https://doi.org/10.3390/rs11232765

Nickel, P. J., Kudina, O., & van de Poel, I. (2022). Moral uncertainty in technomoral change: Bridging the explanatory gap. *Perspectives on Science*, *30*(2), 260–283. https://doi.org/10.1162/posc_a_00414

Nirav Shah, M., & Ganatra, A. (2022). A systematic literature review and existing challenges toward fake news detection models. *Social Network Analysis and Mining*, *12*(1), 168. https://doi.org/10.1007/s13278-022-00995-5

Niu, S., Lu, Z., Zhang, A. X., Cai, J., Griggio, C. F., & Heuer, H. (2023). Building credibility, trust, and safety on video-sharing platforms. *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3544549.3573809

Nix, M., Onisiforou, G., & Painter, A. (2022). Understanding healthcare workers' confidence in AI. *National Health Service, UK*. Retrieved from https://digital-transformation.hee.nhs.uk/binaries/content/assets/digital-transformation/dart-ed/developingconfidenceinai-oct2022.pdf

Nothwang, W. D., McCourt, M. J., Robinson, R. M., Burden, S. A., & Curtis, J. W. (2016). The human should be part of the control loop? *2016 Resilience Week (RWS)*, 214–220. https://doi.org/10.1109/RWEEK.2016.7573336

Nunavath, V., & Goodwin, M. (2019). The use of artificial intelligence in disaster management: A systematic literature review. *2019 International Conference on Information and Communication Technologies for Disaster Management (ICT-DM)*, 1–8. https://doi.org/10.1109/ICT-DM47966.2019.9032935

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, *32*(2), 303–330. https://doi.org/10.1007/s11109-010-9112-2

O'Brien, S. (2019). Translation technology and disaster management. In *The Routledge handbook of translation and technology* (pp. 304–318). Routledge. https://doi.org/10.4324/9781315311258-18

OCHA data responsibility guidelines – The Centre for Humanitarian Data. (n.d.). Retrieved 2 September 2025, from https://centre.humdata.org/the-ocha-data-responsibility-guidelines/

O'Donnell, J., & Crownhart, C. (n.d.). We did the math on AI's energy footprint. Here's the story you haven't heard. *MIT Technology Review*. Retrieved 2 September 2025, from https://www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/

Odubola, O., Adeyemi, T. S., Olajuwon, O. O., Iduwet, N. P., Aniekan, A. I., & Odubola, T. (2025). AI in social good: LLM powered interventions in crisis management and disaster response. *Journal of Artificial Intelligence, Machine Learning and Data Science*, *3*(1), 3353–3360. https://doi.org/10.51219/JAIMLD/Oluwatimilehin-Odubola/510

Oketunji, A. F., Anas, M., & Saina, D. (2023). Large Language Model (LLM) Bias Index: LLMBI. *arXiv Preprint arXiv:2312.14769*. https://doi.org/10.48550/arXiv.2312.14769

Ozmen Garibay, O., Winslow, B., Andolina, S., Antona, M., Bodenschatz, A., Coursaris, C.,…Grieman, K. (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

Palen, L., & Anderson, K. M. (2016). Crisis informatics: New data for extraordinary times. *Science*, *353*(6296), 224–225. https://doi.org/10.1126/science.aag2579

Pan, X., Yang, T. T., Li, J., Ventura, C., Malaga-Chuquitaype, C., Li, C.,…Brzev, S. (2025). A review of recent advances in data-driven computer vision methods for structural damage evaluation: Algorithms, applications, challenges, and future opportunities. *Archives of Computational Methods in Engineering*, 1–33. https://doi.org/10.1007/s11831-025-10279-8

Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, *39*(2), 230–253. https://doi.org/10.1518/001872097778543886

Paulus, D., Fathi, R., Fiedrich, F., de Walle, B. V., & Comes, T. (2024). On the interplay of data and cognitive bias in crisis information management: An exploratory study on epidemic response. *Information Systems Frontiers*, *26*(2), 391–415. https://doi.org/10.1007/s10796-022-10241-0

Pelrine, K., Imouza, A., Thibault, C., Reksoprodjo, M., Gupta, C., Christoph, J.,…Rabbany, R. (2023). Towards reliable misinformation mitigation: Generalization, uncertainty, and GPT-4. *arXiv Preprint arXiv:2305.14928*. https://doi.org/10.18653/v1/2023.emnlp-main.395

Piccolo, L. S., Roberts, S., Iosif, A., & Alani, H. (2018). Designing chatbots for crises: A case study contrasting potential and reality. *Proceedings of the 32nd International BCS Human Computer Interaction Conference*. https://doi.org/10.14236/ewic/HCI2018.56

Pota, M., Pecoraro, G., Rianna, G., Reder, A., Calvello, M., & Esposito, M. (2022). Machine learning for the definition of landslide alert models: A case study in Campania region, Italy. *Discover Artificial Intelligence*, *2*(1), 15. https://doi.org/10.1007/s44163-022-00033-5

Pretolesi, D., Zechner, O., Guirao, D. G., Schrom-Feiertag, H., & Tscheligi, M. (2023). AI-supported XR training: Personalizing medical first responder training. *International Conference on Artificial Intelligence and Virtual Reality*, 343–356. https://doi.org/10.1007/978-981-99-9018-4_25

Qadir, J., Ali, A., ur Rasool, R., Zwitter, A., Sathiaseelan, A., & Crowcroft, J. (2016). Crisis analytics: Big data-driven crisis response. *Journal of International Humanitarian Action*, *1*(1), 12. https://doi.org/10.1186/s41018-016-0013-9

Quach, K. (2020). *AI me to the Moon… Carbon footprint for 'training GPT-3' same as driving to our natural satellite and back*. The Register. Retrieved from https://www.theregister.com/2020/11/04/gpt3_carbon_footprint_estimate/

Rahwan, I., Cebrian, M., Obradovich, N., Bongard, J., Bonnefon, J.-F., Breazeal, C.,…Jackson, M. O. (2019). Machine behaviour. *Nature*, *568*(7753), 477–486. https://doi.org/10.1038/s41586-019-1138-y

Ramchurn, S. D., Huynh, T. D., Wu, F., Ikuno, Y., Flann, J., Moreau, L.,…Simpson, E. (2016). A disaster response system based on human-agent collectives. *Journal of Artificial Intelligence Research*, *57*, 661–708. https://doi.org/10.1613/jair.5098

Rasool, T., Butt, W. H., Shaukat, A., & Akram, M. U. (2019). Multi-label fake news detection using multi-layered supervised learning. *Proceedings of the 2019 11th International Conference on Computer and Automation Engineering*, 73–77. https://doi.org/10.1145/3313991.3314008

Ray, E. C., Merle, P. F., & Lane, K. (2025). Generating credibility in crisis: Will an AI-scripted response be accepted? *International Journal of Strategic Communication*, *19*(2), 158–175. https://doi.org/10.1080/1553118X.2024.2435494

Raymond, N., & Al Achkar, Z. (2016). Data preparedness: Connecting data, decision-making and humanitarian response. *Harvard Humanitarian Initiative: Signal Program on Human Security and Technology-Standards and Ethics Series*, *1*. Retrieved from https://hhi.harvard.edu/sites/g/files/omnuum6866/files/humanitarianinitiative/files/data_preparedness_update.pdf

Raymond, N., Al Achkar, Z., Verhulst, S., Berens, J., Barajas, L., & Easton, M. (2016). Building data responsibility into humanitarian action. *OCHA Policy and Studies Series.* Retrieved from https://ssrn.com/abstract=3141479

Recital 46: Vital interests of the data subject. *General Data Protection Regulation (GDPR)*. Retrieved 2 September 2025, from https://gdpr-info.eu/recitals/no-46/

Reichstein, M., Benson, V., Blunk, J., Camps-Valls, G., Creutzig, F., Fearnley, C. J.,…Schölkopf, B. (2025). Early warning of complex climate risk with integrated artificial intelligence. *Nature Communications*, *16*(1), 2564. https://doi.org/10.1038/s41467-025-57640-w

Rescue Global. (n.d.). Rescue Global. Retrieved 2 September 2025, from https://www.rescueglobal.org

Responsible AI use can advance risk communication and infodemic management in emergencies, new study shows. (n.d.). Retrieved 2 September 2025, from https://www.who.int/europe/news/item/23-05-2025-responsible-ai-use-can-advance-risk-communication-and-infodemic-management-in-emergencies--new-study-shows

Reuel, A., Hardy, A., Smith, C., Lamparth, M., Hardy, M., & Kochenderfer, M. J. (2024). Betterbench: Assessing AI benchmarks, uncovering issues, and establishing best practices. *Advances in Neural Information Processing Systems*, *37*, 21763–21813. https://doi.org/10.52202/079017-0685

Reuter, C., & Kaufhold, M.-A. (2018). Fifteen years of social media in emergencies: A retrospective review and future directions for crisis informatics. *Journal of Contingencies and Crisis Management*, *26*(1), 41–57. https://doi.org/10.1111/1468-5973.12196

Riebe, T., Kaufhold, M.-A., & Reuter, C. (2021). The impact of organizational structure and technology use on collaborative practices in computer emergency response teams: An empirical study. *Proceedings of the ACM on Human-Computer Interaction*, *5*(CSCW2), 1–30. https://doi.org/10.1145/3479865

Röösli, E., Rice, B., & Hernandez-Boussard, T. (2021). Bias at warp speed: How AI may contribute to the disparities gap in the time of COVID-19. *Journal of the American Medical Informatics Association*, *28*(1), 190–192. https://doi.org/10.1093/jamia/ocaa210

Roozenbeek, J., Culloty, E., & Suiter, J. (2023). Countering misinformation. *European Psychologist*. https://doi.org/10.1027/1016-9040/a000492

Rosenthal, U., Charles, M. T., Hart, P. T., Kouzmin, A., & Jarman, A. (1989). From case studies to theory and recommendations: A concluding analysis. *In U. Rosenthal, M. T. Charles, & P. T. Hart (Eds.), Coping with Crises: The Management of Disasters, Riots and Terrorism.*, 436-472. Retrieved from https://psycnet.apa.org/record/1989-98322-017

Sandvik, K. B., Jumbert, M. G., Karlsrud, J., & Kaufmann, M. (2014). Humanitarian technology: A critical research agenda. *International Review of the Red Cross*, *96*(893), 219–242. https://doi.org/10.1017/S1816383114000344

Santoni de Sio, F., & Van den Hoven, J. (2018). Meaningful human control over autonomous systems: A philosophical account. *Frontiers in Robotics and AI*, *5*, 323836. https://doi.org/10.3389/frobt.2018.00015

SAPEA. (2022). *Strategic crisis management in the EU: Evidence review report*. https://doi.org/10.26356/crisismanagement

Savelberg, L., Casali, Y., van den Homberg, M., Zatarain Salazar, J., & Comes, T. (2025). Comparing hierarchical and inductive methods reveals fundamental differences in social vulnerability rankings. *Scientific Reports*, *15*(1), 34541. https://doi.org/10.1038/s41598-025-17860-y

Schmid, S., Hartwig, K., Cieslinski, R., & Reuter, C. (2024). Digital resilience in dealing with misinformation on social media during COVID-19: A web application to assist users in crises. *Information Systems Frontiers*, *26*(2), 477–499. https://doi.org/10.1007/s10796-022-10347-5

Schwartz, R., Dodge, J., Smith, N. A., & Etzioni, O. (2020). Green AI. *Communications of the ACM*, *63*(12), 54–63. https://doi.org/10.1145/3381831

References

Scolobig, A., Potter, S., Kox, T., Kaltenberger, R., Weyrich, P., Chasco, J.,…Uprety, D. (2022). Connecting warning with decision and action: A partnership of communicators and users. In *Towards the "perfect" weather warning: Bridging disciplinary gaps through partnership and communication,* 47–85. Springer International Publishing Cham. https://doi.org/10.1007/978-3-030-98989-7_3

See, L., Chen, Q., Crooks, A., Bayas, J. C. L., Fraisl, D., Fritz, S.,…Lesiv, M. (2025). New directions in mapping the Earth's surface with citizen science and generative AI. *iScience*, *28*(3). https://doi.org/10.1016/j.isci.2025.111919

Sendai Framework for Disaster Risk Reduction 2015-2030 | UNDRR. (2015, June 29). Retrieved from https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030

Seppänen, H., & Virrantaus, K. (2015). Shared situational awareness and information quality in disaster management. *Safety Science*, *77*, 112–122. https://doi.org/10.1016/j.ssci.2015.03.018

Sharma, N., Liao, Q. V., & Xiao, Z. (2024). Generative echo chamber? Effect of LLM-powered search systems on diverse information seeking. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1–17. https://doi.org/10.1145/3613904.3642459

Sharon, T. (2021). Blind-sided by privacy? Digital contact tracing, the Apple/Google API and big tech's newfound role as global health policy makers. *Ethics and Information Technology*, *23* (Suppl 1), 45–57. https://doi.org/10.1007/s10676-020-09547-x

Sherman, I. N., Stokes, J. W., & Redmiles, E. M. (2021). Designing media provenance indicators to combat fake media. *Proceedings of the 24th International Symposium on Research in Attacks, Intrusions and Defenses*, 324–339. https://doi.org/10.1145/3471621.3471860

Shneiderman, B. (2020). Human-centered artificial intelligence: Reliable, safe & trustworthy. *International Journal of Human–Computer Interaction*, *36*(6), 495–504. https://doi.org/10.1080/10447318.2020.1741118

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, *631*(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Siffels, L. E. (2021). Beyond privacy vs. health: A justification analysis of the contact-tracing apps debate in the Netherlands. *Ethics and Information Technology*, *23* (Suppl 1), 99–103.

Sirenko, M., Comes, T., & Verbraeck, A. (2025). The rhythm of risk: Exploring spatio-temporal patterns of urban vulnerability with ambulance calls data. *Environment and Planning B: Urban Analytics and City Science*, *52*(4), 863–881. https://doi.org/10.1177/23998083241272095

Stauffer, M., Seifert, K., Aristizábal, A., Chaudhry, H. T., Kohler, K., Hussein, S. N.,…Estier, M. (2023). *Existential risk and rapid technological change: Advancing risk-informed development*. United Nations Office for Disaster Risk Reduction Geneva, Switzerland. Retrieved from https://www.undrr.org/media/86500/download?startDownload=20251119

Sun, Y. Q., Hassanzadeh, P., Zand, M., Chattopadhyay, A., Weare, J., & Abbot, D. S. (2025). Can AI weather models predict out-of-distribution gray swan tropical cyclones? *Proceedings of the National Academy of Sciences*, *122*(21), e2420914122. https://doi.org/10.1073/pnas.2420914122

't Hart, P., Rosenthal, U., & Kouzmin, A. (1993). Crisis decision making: The centralization thesis revisited. *Administration & Society*, *25*(1), 12–45. https://doi.org/10.1177/009539979302500102

Takabatake, T., Asai, K., Kakuta, H., & Hasegawa, N. (2025). Optimizing evacuation paths using agent-based evacuation simulations and reinforcement learning. *International Journal of Disaster Risk Reduction*, *117*, 105173. https://doi.org/10.1016/j.ijdrr.2024.105173

Tan, J., Sumpena, E., Zhuo, W., Zhao, Z., Liu, M., & Chan, S.-H. G. (2020). IoT geofencing for COVID-19 home quarantine enforcement. *IEEE Internet of Things Magazine*, *3*(3), 24–29. https://doi.org/10.1109/IOTM.0001.2000097

Tanti, L., Efendi, S., Lydia, M. S., & Mawengkang, H. (2022). Designing an optimization model for dynamic facility-location problem at post-disaster area considering uncertainty. *2022 4th International Conference on Cybernetics and Intelligent System (ICORIS)*, 1–6. https://doi.org/10.1109/ICORIS56080.2022.10031495

Tarkhanova, O. (2024). The politics of (im) mobility: The effects of the pandemic on movement across the 'contact line' in Eastern Ukraine. *Europe-Asia Studies*, *76*(9), 1392–1416. https://doi.org/10.1080/09668136.2023.2281239

Tiggeloven, T., Pfeiffer, S., Matanó, A., van den Homberg, M., Thalheimer, L., Reichstein, M., & Torresan, S. (2025). The role of artificial intelligence for early warning systems: Status, applicability, guardrails and ways forward. *iScience*. https://doi.org/10.1016/j.isci.2025.113689

Tondaś, D., Kazmierski, K., & Kapłon, J. (2023). Real-time and near real-time displacement monitoring with GNSS observations in the mining activity areas. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *16*, 5963–5972. https://doi.org/10.1109/JSTARS.2023.3290673

Trzun, Z. (2024). Artificial intelligence and human-out-of-the-loop: Is it time for autonomous military systems? *European Integration Studies*, *20*(2), 479–508. https://doi.org/10.46941/2024.2.18

Turoff, M., Chumer, M., de Walle, B. V., & Yao, X. (2004). The design of a dynamic emergency response management information system (DERMIS). *Journal of Information Technology Theory and Application (JITTA)*, *5*(4), 3. Retrieved from https://www.researchgate.net/publication/308162811_The_design_of_a_dynamic_emergency_response_management_information_system

Unmanned aerial vehicles in humanitarian response | OCHA. (2014, June 30). Retrieved from https://www.unocha.org/publications/report/world/unmanned-aerial-vehicles-humanitarian-response

Urbanelli, A., Frisiello, A., Bruno, L., & Rossi, C. (2024). The ERMES chatbot: A conversational communication tool for improved emergency management and disaster risk reduction. *International Journal of Disaster Risk Reduction*, *112*, 104792. https://doi.org/10.1016/j.ijdrr.2024.104792

Vallor, S. (2015). Moral deskilling and upskilling in a new machine age: Reflections on the ambiguous future of character. *Philosophy & Technology*, *28*(1), 107–124. https://doi.org/10.1007/s13347-014-0156-9

van Brakel, R., Kudina, O., Fonio, C., & Boersma, K. (2022). Bridging values: Finding a balance between privacy and control. The case of Corona apps in Belgium and the Netherlands. *Journal of Contingencies and Crisis Management*, *30*(1), 50–58. https://doi.org/10.1111/1468-5973.12395

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*, *95*, 66–79. https://doi.org/10.1016/j.ijhcs.2016.05.001

Van de Walle, B., & Comes, T. (2015). On the nature of information management in complex and natural disasters. *Procedia Engineering*, *107*, 403–411. https://doi.org/10.1016/j.proeng.2015.06.098

Van Den Homberg, M., Visser, J., & Van Der Veen, M. (2017). Unpacking data preparedness from a humanitarian decision-making perspective: Toward an assessment framework at subnational level. *Proc. Information Systems for Crisis Response and Management (ISCRAM) Conference*. https://ris.utwente.nl/ws/portalfiles/portal/499575503/1995_MarcvandenHomberg_etal2017.pdf

van Leersum, C. M., & Maathuis, C. (2025). Human centred explainable AI decision-making in healthcare. *Journal of Responsible Technology*, *21*, 100108. https://doi.org/10.1016/j.jrt.2025.100108

van Leeuwen, B., Gasaway, R., Spaling, G., & Netage, B. V. (2022). Adopting AI to support situational awareness in emergency response: A reflection by professionals. *Proceedings of the First International Conference on Hybrid Human-Artificial Intelligence. Frontiers in Artificial Intelligence and Applications*. Retrieved from https://www.hhai-conference.org/wp-content/uploads/2022/08/hhai-2021_paper_96.pdf

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. https://doi.org/10.1007/s10676-019-09514-1

Van Wynsberghe, A., & Robbins, S. (2019). Critiquing the reasons for making artificial moral agents. *Science and Engineering Ethics*, *25*(3), 719–735. https://doi.org/10.1007/s11948-018-0030-8

Vanjani, M., Aiken, M., & Park, M. (2019). Chatbots for multilingual conversations. *Journal of Management Science and Business Intelligence*, *4*(1), 19–24. https://southernct.elsevierpure.com/en/publications/chatbots-for-multilingual-conversations-2/

Vargas, L., Emami, P., & Traynor, P. (2020). On the detection of disinformation campaign activity with network analysis. *Proceedings of the 2020 ACM SIGSAC Conference on Cloud Computing Security Workshop*, 133–146. https://doi.org/10.1145/3411495.3421363

Vassell, M., Apperson, O., Calyam, P., Gillis, J., & Ahmad, S. (2016). Intelligent dashboard for augmented reality-based incident command response co-ordination. *2016 13th IEEE Annual Consumer Communications & Networking Conference (CCNC)*, 976–979. https://doi.org/10.1109/CCNC.2016.7444921

Venkataramanan, A., Bodesheim, P., & Denzler, J. (2025). Probabilistic embeddings for frozen vision-language models: Uncertainty quantification with Gaussian Process Latent Variable Models. *arXiv Preprint arXiv:2505.05163*. https://doi.org/10.48550/arXiv.2505.05163

Verbeek, P. P., Brey, P., Van Est, R., van Gemert, L., Heldeweg, M., & Moerel, L. (2020). *Ethical analysis of the COVID-19 notification app to supplement epidemiological source and contact research*. Retrieved from https://research.tue.nl/files/165848982/ethische_analyse_van_de_covid_19_notificatie_app_ter_aanvulling_op_bron_en_contactonderzoek_ggd.pdf

Vinnell, L. J., Becker, J. S., Scolobig, A., Johnston, D. M., Tan, M. L., & McLaren, L. (2021). Citizen science initiatives in high-impact weather and disaster risk reduction. *Australasian Journal of Disaster and Trauma Studies*, *25*(3), 55–60.

Visave, J. (2025). Transparency in AI for emergency management: Building trust and accountability. *AI and Ethics*, 1–14. https://doi.org/10.1007/s43681-025-00692-x

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. https://doi.org/10.1109/TGRS.2007.895830

Wagenmakers, E.-J., Sarafoglou, A., & Aczel, B. (2022). One statistical analysis must not rule them all. *Nature*, *605*(7910), 423–425. https://doi.org/10.1038/d41586-022-01332-8.

References

Wang, H., Hee, M. S., Awal, M. R., Choo, K. T. W., & Lee, R. K.-W. (2023). Evaluating GPT-3 generated explanations for hateful content moderation. *arXiv Preprint arXiv:2305.17680*. https://doi.org/10.24963/ijcai.2023/694

Wankmüller, C., Kunovjanek, M., & Mayrgündter, S. (2021). Drones in emergency response–evidence from cross-border, multi-disciplinary usability tests. *International Journal of Disaster Risk Reduction*, *65*, 102567. https://doi.org/10.1016/j.ijdrr.2021.102567

Weick, K. E., & Weick, K. E. (1995). *Sensemaking in organizations* (Vol. 3, Issue 10.1002). Sage publications Thousand Oaks, CA.

Weik, K. E. (2009). The collapse of sensemaking in organizations: The Mann Gulch disaster. *Studi Organizzativi*, *2008/2*. https://doi.org/10.3280/SO2008-002009

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A.,… Bourne, P. E. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 1–9. https://doi.org/10.1038/sdata.2016.18

Williams, J. C., Anderson, N., Mathis, M., Sanford III, E., Eugene, J., & Isom, J. (2020). Colorblind algorithms: Racism in the era of COVID-19. *Journal of the National Medical Association*, *112*(5), 550–552. https://doi.org/10.1016/j.jnma.2020.05.010

Winiecki, D., Spezzano, F., & Underwood, C. (2023). Understanding teenagers' real and fake news sharing on social media. *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference*, 598–602. https://doi.org/10.1145/3585088.3593864

Wood, T., & Porter, E. (2019). The elusive backfire effect: Mass attitudes' steadfast factual adherence. *Political Behavior*, *41*(1), 135–163. https://doi.org/10.1007/s11109-018-9443-y

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, *135*, 364–381. https://doi.org/10.1016/j.future.2022.05.014

Wynants, L., Van Calster, B., Collins, G. S., Riley, R. D., Heinze, G., Schuit, E.,…Bonten, M. M. (2020). Prediction models for diagnosis and prognosis of COVID-19: Systematic review and critical appraisal. *Bmj*, *369*.

Xu, Y., Lu, X., Cetiner, B., & Taciroglu, E. (2021). Real-time regional seismic damage assessment framework based on long short-term memory neural network. *Computer-Aided Civil and Infrastructure Engineering*, *36*(4), 504–521. https://doi.org/10.1111/mice.12628

Yabe, T., Rao, P. S. C., Ukkusuri, S. V., & Cutter, S. L. (2022). Toward data-driven, dynamical complex systems approaches to disaster resilience. *Proceedings of the National Academy of Sciences*, *119*(8), e2111997119. https://doi.org/10.1073/pnas.2111997119

Yang, C. (2022). Digital contact tracing in the pandemic cities: Problematizing the regime of traceability in South Korea. *Big Data & Society*, *9*(1), 20539517221089294. https://doi.org/10.1177/20539517221089294

Yang, F., Heemsbergen, L., & Fordyce, R. (2021). Comparative analysis of China's Health Code, Australia's COVIDSafe and New Zealand's COVID Tracer Surveillance apps: A new corona of public health governmentality? *Media International Australia*, *178*(1), 182–197.

Yao, H., Rashidian, S., Dong, X., Duanmu, H., Rosenthal, R. N., & Wang, F. (2020). Detection of suicidality among opioid users on Reddit: Machine learning–based approach. *Journal of Medical Internet Research*, *22*(11), e15293. https://doi.org/10.2196/15293

Yokoyama, H., & Takefuji, Y. (2026). Unbiased evaluation of social vulnerability: A multimethod approach using machine learning and nonparametric statistics. *Cities*, *168*, 106519. https://doi.org/10.1016/j.cities.2025.106519

Zamfirescu-Pereira, J. D., Wong, R. Y., Hartmann, B., & Yang, Q. (2023). Why Johnny can't prompt: How non-AI experts try (and fail) to design LLM prompts. *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–21. https://doi.org/10.1145/3544548.3581388

Zhang, C., Liu, X., Jiang, Y. P., Fan, B., & Song, X. (2016). A two-stage resource allocation model for lifeline systems quick response with vulnerability analysis. *European Journal of Operational Research*, *250*(3), 855–864. https://doi.org/10.1016/j.ejor.2015.10.022

Zhang, T., Wang, D., & Lu, Y. (2023). Machine learning-enabled regional multi-hazards risk assessment considering social vulnerability. *Scientific Reports*, *13*(1), 13405. https://doi.org/10.1038/s41598-023-40159-9

Zhou, X., Wu, J., & Zafarani, R. (2020). SAFE: Similarity-aware multi-modal fake news detection. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 354–367. https://doi.org/10.1007/978-3-030-47436-2_27

Zhou, Y., Wu, Z., Jiang, M., Xu, H., Yan, D., Wang, H.,…Zhang, X. (2024). Real-time prediction and ponding process early warning method at urban flood points based on different deep learning methods. *Journal of Flood Risk Management*, *17*(1), e12964. https://doi.org/10.1111/jfr3.12964

Zwitter, A. (2018). Principles and professionalism: Towards humanitarian intelligence. In H.-J. Heintze & P. Thielbörger (Eds.), *International Humanitarian Action: NOHA Textbook* (pp. 103–120). Springer International Publishing. https://doi.org/10.1007/978-3-319-14454-2_6

Zwitter, A., & Gstrein, O. J. (2020). Big data, privacy and COVID-19 – learning from humanitarian expertise in data protection. *Journal of International Humanitarian Action*, *5*(1), 4. https://doi.org/10.1186/s41018-020-00072-6

Annexes

Annex 1. Background

Scientific Advice Mechanism

The **Scientific Advice Mechanism** provides independent scientific evidence and policy recommendations to the College of European Commissioners on any subject, including on policy issues that the European Parliament and the Council consider to be of major importance.

It consists of three parts:

- The **Group of Chief Scientific Advisors**, seven eminent scientists whose role is to make policy recommendations
- **SAPEA** (Science Advice for Policy by European Academies), which brings together Europe's academies and Academy Networks to review and synthesise evidence
- The **SAM secretariat**, a unit within the European Commission whose role is to support the Advisors and liaise between the Scientific Advice Mechanism and the European Commission

When giving scientific advice, SAPEA and the Advisors work independently, following specific procedures to maintain the independence and quality of the SAM's advice.

The request

The **Emergency Response Coordination Centre** (ERCC), which is the core of the EU Civil Protection Mechanism, has tasked the Scientific Advice Mechanism with producing an evidence review report. The overarching questions posed in the Specifications of Work were:

Based on the evidence, what are the characteristics, opportunities and risks associated with the use of artificial intelligence in crisis preparedness and response? According to the literature, how can these risks be mitigated?

SAPEA was tasked to deliver the evidence review report, including evidence-informed conclusions and options for policy.

The Specifications of Work were co-developed by the ERCC and SAPEA, with support from the SAM Secretariat. They were finalised and approved in July 2025, with a content-final evidence review report due in November 2025, and the final publication expected by December 2025.

Annex 2. Process

When the deadline set by the European Commission to prepare an evidence review report is very short, SAPEA can prepare a Rapid Evidence Review Report, as described in SAPEA's Quality Assurance Guidelines[53]. To respond to the request while keeping to the timeline, the guidelines were implemented as follows.

Selection of experts for the Working Group.
SAPEA established a small interdisciplinary working group consisting of 8 experts. The Chair, Professor Tina Comes, was nominated by ALLEA, the lead Academy Network. The Chair was approved by the SAPEA Board, following assessment of her declaration of interest.

The Chair helped define the necessary areas of expertise to respond to the request. Suggestions for experts were then invited through the Academy Networks of SAPEA. Together, the Networks suggested a total of nearly 70 experts. From this pool, a final Working Group of 7 experts (plus Chair) was selected, based primarily on demonstrated excellence in the predefined fields of expertise.

The composition of the Working Group was approved by the SAPEA Board. All Working Group members were required to complete the standard Declaration of Interests form of the European Commission, which was then assessed by SAPEA in accordance with SAPEA's Quality Guidelines. In the assessment, no conflicts of interests were detected.

Following feedback from the expert workshop (see Expert Workshop Report), the Working Group Chair decided to include an additional case study to reflect on the use of AI during the COVID-19 pandemic. An additional expert was invited as a contributor to the Rapid Evidence Review Report.

Working Group
The Working Group met three times during the period from July 2025 to November 2025, via online meetings. Between meetings, they worked collectively online on successive drafts of the Report.

Literature review
While the scope and process were still being defined, it was decided to conduct an initial literature search on the topic. The purpose of the search was to provide a complementary evidence base, and it was conducted by Cardiff University. The first draft of the narrative review was produced in May 2025. It was revised iteratively, in response to feedback from the SAM Secretariat, DG ECHO and the Working Group. The final version is published separately to the Rapid Evidence Review Report and is available on the SAM website.

Expert workshop
An expert workshop was held online on 6th October 2025. Its purpose was to receive feedback on the

53 https://scientificadvice.eu/reports/quality-assurance-guidelines-and-procedures-on-science-advice-for-policy-and-society/

draft Rapid Evidence Review Report from the wider expert community. The workshop served as a peer-review step, as experts were also invited to provide written feedback ahead of the workshop.

To identify experts, a call for nominations was sent to all Academy Networks and member Academies. Experts were selected based on their expertise, while also considering the diversity and inclusiveness criteria set out in SAPEA's strategy on EDI. The selection was carried out by SAPEA Scientific Policy Officers, with guidance from the Working Group Chair. The pool of experts to be invited was approved by the SAPEA Board. In the final group of 17 invited experts, 41% of experts were female; 47% early- and mid-career researchers. 10 European countries of work were represented in the group.

At the workshop, there were 54 participants, including the invited experts, SAPEA Working Group, Members of the Group of Chief Scientific Advisors to the European Commission, European Commission representatives, SAPEA and SAM Secretariat staff.

The workshop started with a general overview of the Rapid Evidence Review Report. It presented the Report's strengths and identified gaps, based on the written feedback, followed by a facilitated discussion on these issues. The main points of feedback were considered by the Working Group at a dedicated meeting. The complete set of written comments was also discussed separately and addressed after the workshop.

The report of the workshop is published separately, as a companion document to the Rapid Evidence Review Report, and is available on the SAM website. The invited experts are listed in Annex 3.

Revisions following the expert workshop
The Working Group considered all the written and oral comments received at the workshop. They addressed them, keeping in mind the scope, nature and timeline for the Rapid Evidence Review Report. The main actions taken per chapter included the following:

Clarity, structure, methods
Sections were more tightly connected and integrated. A preface was added to clarify the scope of the Report, and annexes explaining the process were included.

Definitions and framing
The section was reworked to improve synthesis and integration. Terminology was clarified, figures were added to illustrate the framework, and the intended uses and purposes of AI tools in emergency and crisis management were made clearer.

Performance of AI across tasks: Reflections on task allocation and control were added.

- For monitoring, predicting, and anticipating: more detail and examples of AI performance were included, as well as highlighting ensemble approaches to improve forecasting.
- For assessing and reporting: reflections on the role of LLMs and evidence on assessing resilience and vulnerability were added.

- For decision-support: information was added on control frameworks (Human-in-the-loop/Human-on-the-loop/Human-out-of-the-loop). Additional content clarified where AI should not be used, particularly in relation to meaningful human control.

Legal and ethical requirements and guiding principles

Additional detail was provided on the classification of "high-risk" AI systems and on the use of general-purpose AI in crisis management. A new section on the AI Act in practice was added, including gaps, requirements, and operational implications, along with further elaboration of the "Do No Harm" principle. A section was added on predefined decision-making rules and accountability under uncertainty.

Data governance and sharing

Considerations on data quality were added, along with a box highlighting Europe's critical dependencies on external data infrastructures.

Data preparedness

Reflections were added on the potential for AI to support data-sparse contexts, along with references to complementary data-sharing frameworks.

User uptake

Additional evidence and considerations on human–AI teaming were included.

Case studies and examples

The case study structure was further harmonised, lessons learned were added, and a case study on the use of AI during the COVID-19 pandemic was included.

Plagiarism check

In accordance with the SAPEA Quality Guidelines, a plagiarism check on the final version of the evidence review report was run by Cardiff University, using Turnitin software. The results were checked with the Scientific Policy Officer of ALLEA.

Publication

This *Rapid Evidence Review Report* has been published and handed over on 11 December 2025. It is accompanied by the *Expert Workshop Report* and an introductory narrative review of recent literature (*Literature Review*). All documents can be accessed on the SAM website[54].

54 http://scientificadvice.eu

Annex 3. Acknowledgements

SAPEA wishes to thank the following people for their valued contributions and support in the production of this report.

Working Group and Contributors

The Working Group members who wrote this report are listed at the start.

Expert workshop participants

- **Burcu Balçık,** Özyeğin University
- **Maurice de Beer,** Rotterdam-Rijnmond Fire Service
- **Marc Berenguer Ferrer,** Universitat Politècnica de Catalunya, Barcelona
- **Quentin Brot,** The French National Fire Officers Academy (ENSOSP)
- **Mario Drobics,** AIT Austrian Institute of Technology
- **Adina Florea,** National University of Science and Technology POLITEHNICA Bucharest (UNSTPB)
- **Oskar Josef Gstrein,** University of Groningen
- **Marc van den Homberg,** University of Twente
- **Kai Kornhuber,** International Institute for Applied Systems Analysis (IIASA)
- **Olya Kudina,** TU Delft
- **Qiuhua Liang,** Loughborough University
- **Igor Linkov,** US Army Corps of Engineers and Adjunct Professor at Carnagie Mellon University
- **Marta López-Saavedra,** Natural Risks Assessment and Management Service (NRAMS), Institute of Environmental Diagnosis and Water Research (IDAEA-CSIC), Spanish National Research Council (CSIC)
- **Warner Marzocchi,** University of Naples Federico II
- **Shruti Nath,** University of Oxford
- **Karen Yeung,** University of Birmingham
- **Barbara Zitová,** Czech Academy of Sciences

SAPEA staff

- **Céline Tschirhart,** ALLEA Scientific Policy Officer (lead)
- **Louise Edwards,** AE Scientific Policy Officer, Cardiff University
- **Alice Sadler,** Cardiff University
- **Sara Saba,** ALLEA Scientific Policy Officer
- **Lionel Dutrieux,** Digital Communications Officer

- **Justine Moynat,** Communications Manager
- **Alaa Jbour,** Head of Communications
- **Anda Popovici,** YASAS Scientific Policy Officer
- **Frederico Rocha,** EASAC Scientific Policy Officer
- **Carly Seedall,** AE Scientific Policy Officer
- **Mariana Rei,** Euro-CASE Scientific Policy Officer
- **Rudolf Hielscher,** SAPEA Manager

Science Policy, Advice and Ethics Unit at DG RTD, European Commission

- **Jonathan Murphy,** Policy Officer
- **Geoffroy Delamare,** Policy Officer
- **Ingrid Zegers,** Team Lead
- **Annabelle Ascher,** Secretary to the Group of Chief Scientific Advisors

scientificadvice.eu

Within the Scientific Advice Mechanism, SAPEA *is funded by the European Union (grant number 101198044).*

Funded by the European Union