Death by AI: A Survey of the Literature and Known Incidents

Lauri Tuovinen

Biomimetics and Intelligent Systems Group, P.O. Box 4500, FI-90014 University of Oulu, Finland

Abstract

When the operation of an artificial intelligence (AI) system results in loss of life, this is arguably the ultimate failure of AI safety. To better understand how this may occur, a search of fatal or near-fatal incidents was carried out in three public AI incident databases. From the results, five major incident categories were identified: robots, essential services, hazardous guidance, algorithmic recommendations and conversational agents. These were compared and contrasted with themes identified in a survey of academic literature on AI, ethics and death. The findings are tentative and require further validation, but they suggest that there are some areas where the interests of academic research are not fully aligned with the requirements of real-world AI safety. Systems that induce violent or suicidal behaviour through subtle psychological influence are a particular concern, as their safety implications may not be as fully understood as those of systems that are directly involved in life-or-death decisions. As AI systems grow increasingly humanlike in their performance and behaviour, cross-disciplinary research efforts are necessary to ensure that they are safe to deploy and interact with.

Keywords

AI ethics, safety, AI incident, autonomous system, automated decisions, recommender system, generative AI

1. Introduction

Stories of artificial intelligence (AI) systems that turn hostile to humans and start killing them have long been a staple of science fiction. Meanwhile, AI systems that exist in the real world have no intentions or emotions and therefore no hostility, but there have still been many documented incidents where an AI system contributed to physical harm, including death. Cyber-physical systems such as autonomous vehicles can cause death directly, but even chatbots, whose sole interface with the physical realm is words on a screen, may cause it indirectly by making harmful suggestions to the user.

When an AI system causes death, it can be considered the ultimate violation of the principle of non-maleficence. Therefore, in the interest of making AI safer, it is instructive to study the fatal incidents that have occurred so far in order to better understand the nature and magnitude of the risk AI systems pose to human life. Previously, AI incidents have been reviewed in [1], but the study was much broader in scope, covering safety and security incidents in general rather than focusing on fatal ones specifically.

Furthermore, it is useful to compare and contrast the incident reports with relevant research literature in order to see how closely academic discourse on AI and death mirrors what is happening in the real world. Previously, academic literature on AI safety has been reviewed in [2], but again, the scope of the study was considerably different. On the one hand, the paper examines AI safety on a much more general level; on the other hand, literature on AI and death also covers topics that are not related to AI safety as such and therefore fall outside the scope of the review.

This paper investigates the intersection of AI, ethics and death using the aforementioned two bodies of material, taking a tentative step toward the systematisation of knowledge in this area. Through qualitative analysis, prominent themes are extracted from the materials to identify potentially noteworthy subtopics and to facilitate comparison of the two domains. To the best of the author's knowledge, no previously published work with this objective exists. By analysing the materials, the following research questions are studied:

8th Conference on Technology Ethics (TETHICS2025), November 11–12, 2025, Vaasa, Finland

lauri.tuovinen@oulu.fi (L. Tuovinen)

0000-0002-7916-0255 (L. Tuovinen)

1. What questions related to AI and death is academic literature concerned with?
2. What characterises known incidents where the operation of an AI system has been either directly or indirectly linked with loss of human life?
3. Where do the two perspectives on AI and death overlap and where do they diverge?
4. What are the implications of the findings in terms of e.g. questions that warrant more attention in AI ethics research?

A more detailed account of the materials and methods used is given in Section 2. The first research question is then addressed in Section 3, which presents a narrative review of relevant literature found on Scopus. Section 4 addresses the second question by classifying (potentially) fatal AI incidents found in three public databases and examining some of the discourse surrounding them in journalistic media. The third and fourth questions are then addressed in Section 5, which discusses the significance and broader implications of the search results. Finally, Section 6 concludes the paper.

2. Methodology and Scope

For the purposes of this paper, AI can be defined as in [3], as ”the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings”. The term is commonly defined in this way, by analogy with organic intelligence and with reference to externally observable behaviour. In terms of technology, AI is used as an umbrella term covering technologies that can be used to implement intelligent behaviour in computer-based systems, including (but not limited to) machine learning.

For the literature review in Section 3, a Scopus search was carried out in March 2025 with the following search string:

> (”artificial intelligence” OR ai) AND ethics AND death

The search results were restricted to journal articles, conference papers and book chapters, and publications in languages other than English were excluded. No publication date range was specified. This resulted in a set of 84 papers, which were examined more closely to identify major themes relevant to the research questions.

To discover AI incidents involving death, three sources were used:

- AI Incident Database (AIID): a repository of incidents that accepts submissions from the general public, with approximately 1000 incidents included at the time of writing [4].
- AI, Algorithmic and Automation Incidents and Controversies (AIAAIC) Repository: similar to AIID, but not limited to incidents involving AI [5].
- OECD AI Incidents and hazards Monitor (AIM): an automatically curated collection of incidents compiled by using AI to classify news articles from various sources [6].

The bulk of the discovered incidents were found by searching AIID and AIAAIC for ”death” and related keywords (”dead”, ”die” and its inflections, ”fatal”, ”kill”, ”murder”, ”suicide”). These were supplemented by using the advanced search of AIM to list incidents with the severity classification ”death”. Besides supplying some incidents not found in either AIID or AIAAIC, AIM provided some insights into the public discourse about certain types of incidents. The discovered incidents were clustered into themes through single-coder inductive analysis following a process similar to the one of Braun and Clarke [7].

Since there is no previous work addressing the specific topic of AI and death, the study reported here was conceived as a tentative exploration of the area and a preamble to more systematic work. The selected data sources were deemed sufficient to provide a reasonably comprehensive picture for this purpose. The methods used were likewise chosen with this purpose in mind: rather than devise a unified methodology for studying both categories of data, it was deemed better to explore the categories separately using relatively lightweight and flexible methods to pave way for a more rigorous approach in the future.

3. Death and AI in the Literature

A quantitative summary of the results of the literature review is presented in Table 1. A large portion of the literature found deals with the use of AI in medicine, and much of this was found to be not particularly interesting in the context of the research questions at hand: AI is viewed as something capable of preventing death rather than causing it, and the discussion of ethics is mainly concerned with affirming that the study was carried out in accordance with the principles of medical research ethics. There are a few notable exceptions, however, the most directly relevant one being [8], which discusses the possibility of a patient's death resulting in manslaughter charges when an AI system has been used for clinical decision-making.

Another interesting paper is [9], where it is argued that prioritising fairness over utility in medical AI may lead to preventable deaths if the deployment of an AI system that performs well on the majority of patients is delayed because of its poor performance on minority groups. Somewhat relevant are also papers that examine the use of AI for potentially life-or-death decisions on a more general level, such as [10] and especially [11], which explicitly proposes the incorporation of moral reasoning into a medical decision support system. Recently, the popular research practice of evaluating the competence of generative AI systems such as ChatGPT in various fields of knowledge has reached the field of medical ethics [12].

In [13], judicial decision-making is brought up as another AI application domain where the outcome of a decision may mean the difference between life and death. The study on bias and discrimination in administrative AI systems presented in [14] can be considered tangentially relevant, as the motivation for the study involves the (non-AI-related) death of an eight-year-old child and the failure of social services to intervene in time. Likewise of tangential relevance is the discussion of recommender systems in [15], which uses the word ”death” figuratively but also mentions some cases where recommendation algorithms have allegedly caused literal death.

A radically different perspective is provided by autonomous AI systems deliberately inflicting harm, including death. In [16], this is referred to as autonomised harming, the ethics of which are discussed in the context of self-driving cars and autonomous weapons. For self-driving cars, [17] presents a broad overview of the ethical issues, while [18] and [19] look more closely at ethical decision-making in dilemma situations where an accident is unavoidable and the vehicle must decide who to harm (and potentially kill). In [20], a study of moral preferences with respect to how self-driving cars should behave in such situations is presented. In [21], it is suggested that such dilemmas can be used as benchmarks in the evaluation of ethical decision-making algorithms. The study of the discourse surrounding robot-related fatalities in [22] covers self-driving cars alongside other autonomous entities such as industrial robots.

For autonomous weapons, the literature provides examples of arguments both for [23] and against [24, 25]. In [26] and [27], the ethical considerations are discussed in a neutral fashion, while [28] emphasises the importance of educating AI developers on ethical issues in ensuring that military AI systems are developed responsibly. In [29], combat drones are juxtaposed with companion robots in a discussion of the capability of AI systems to respect human values and the merits of a utilitarian approach to AI ethics. In [30], a similar juxtaposition is made between robots designed to kill and ones designed to preserve life; the paper points out that even robots in the latter category have been known to cause loss of life and discusses the question of whether morality can be designed into robots.

The recent generative AI boom can be observed also in literature dealing with AI ethics and death. In particular, there has been a surge of interest in so-called deathbots or griefbots: chatbots trained to imitate deceased individuals in the way they interact with the living, usually loved ones of the deceased person. Several authors have explored the ethics of such applications, which are already being offered as a commercial service by a number of companies [31, 32, 33, 34]. Instead of a mourning family member or friend, the initiative could come from the person before their passing or from a company that has collected their data [35]. Another possibility is application of AI to data of the dead for archaeological purposes [36].

The ”resurrected” individual could also be a celebrity. According to the survey reported in [37], there

Table 1
Summary of the Reviewed Literature

Category	Subcategory	Number of papers
Decision support	Medical	5
	Other	3
Autonomised harming	Civilian	6
	Military	6
	Mixed	3
Digital resurrection	Griefbots	5
	Other	3
Other generative AI	Virtual companions	2
	Synthetic media	3
AI as existential risk		3

is currently not much public interest in this, but so-called deepfake technology has been used in the entertainment industry to replicate the likeness of deceased performers, Peter Cushing in the film *Rogue One: A Star Wars Story* being one famous example. A beneficent example of digital resurrection by deepfake is given in [38], which discusses a campaign to raise awareness of intimate partner violence using synthetically generated videos of women who had been murdered by their partners.

Digitally resurrected individuals are but a subset AI systems that simulate a personality and interact socially with their users. Another subset with connections to death is virtual companions; [39] discusses the possibility of these being used to alleviate loneliness and thus to reduce the associated health risks, including the risk of early death. In [40], it is found that people may respond emotionally to the loss of an AI companion with intense grief, not unlike to the death of a loved one.

Journalism is another application domain where synthetically generated content is now having an impact; in this context, [41] raises the question of whether machines should write about death. Another paper from the field of media ethics is [42], which argues that the use of synthetic data is inherently unethical, noting the life-or-death nature of certain types of AI systems such as medical AI. Apart from this, the only paper among the search results to suggest some kind of causal link between synthetic content and death is [43], which discusses the use of speech synthesis to model queer voices and cautions that malicious use of the technology could have fatal consequences for LGBTQ+ individuals.

Despite its status as a common concern as well as an established science fiction trope, AI as an existential risk to humanity is largely absent among the search results. The principal exception is [44], which discusses the long-term risks of AI alongside the potential long-term benefits. Another paper that could be included in this category is [45], which brings up the contribution of AI use to death of the environment. One more that can be considered somewhat relevant is [46], which applies ethical theories to the destruction of infrastructure, although AI is but one of the possible means of destruction mentioned.

4. Death in AI Incidents

Through thematic analysis, five major categories of incidents were identified: robots, essential services, hazardous guidance, algorithmic recommendations and conversational agents. By a considerable margin, the most numerous category was **robots**, with over 50 incidents discovered. This is unsurprising, given that AI systems that operate in the physical realm are the ones with the most immediate safety implications. However, a closer look at the incidents reveals that the high number is to some extent misleading, as a substantial portion of the incidents involve robots that are not controlled by AI. Industrial robots, for instance, are typically controlled by traditional programming, whereas surgical robots are controlled remotely by a human surgeon.

Among incidents involving AI-controlled robots, fatal accidents involving autonomous vehicles are the predominant subcategory, with well over 30 such incidents discovered by the search. The vast majority of these involved various Tesla models specifically; however, in many cases it is unclear whether the Autopilot feature of the vehicle was in fact engaged at the time of the accident. Factors such as the relative popularity of different makes of (partially) self-driving cars must also be accounted for, so the number of incidents as such is not necessarily a good proxy for safety. Nevertheless, there have been enough incidents to cause the National Highway Traffic Safety Administration (NHTSA) of the United States to launch an investigation of the safety of the Full Self-Driving (FSD) feature, the most advanced version of Autopilot [47].

The discourse in the incident reports revolves mainly around safety and liability issues. These are intertwined, since the interface between human and machine introduces complications while it is still at least possible, if not necessary, for a human driver to take over from the AI in an emergency. Concerns have been raised about how this taking over can be done safely [48], as well as how the public may view self-driving cars as more capable than they actually are and how design choices affect this impression [49]. In the case of an accident, the public is more likely to blame the driver than the vehicle according to a study [50]. Regarding Tesla specifically, the company has been criticised for aggressively rolling out unsafe self-driving technology [51] and sending contradictory messages about its capabilities to consumers and regulators [52]. It has also been suspected that the cars have been programmed to disengage Autopilot when a crash is imminent to shift blame away from the company [53].

Some 15 incidents were related to robotic weapons or other military uses of AI. Again, in several cases the robots were remotely controlled with no AI involvement, and in almost all cases the decision to engage was made by a human operator rather than an AI system. Possible exceptions are an incident involving the use of Kargu-2 drones in Libya [54], allegedly the first time a fully autonomous weapon system has been used to attack human targets, and the use of KUB-BLA drones by Russian forces in Ukraine [55]. However, in public discourse even fully operator-controlled robotic weapons may be viewed as controversial, especially when taken out of the military context and adopted by law enforcement organisations [56]. The use of AI-based target identification systems by the Israeli Defence Forces has also sparked controversy [57, 58], although it is difficult to say how much of this should be attributed to the use of AI for this purpose as opposed to the high rate of Palestinian non-combatant casualties in the Gaza conflict.

Incidents involving the use of AI for law enforcement can also be found in the **essential services** category, but here we are dealing with systems that do not directly cause any effects in the physical world. The ShotSpotter gunfire detection system, involved in an incident where a teenager was killed in a police raid [59] is illustrative of a common theme running through the category: systems designed to protect life and well-being that, because of inaccuracy or bias, fail or even do the opposite. Other law enforcement examples include a facial recognition system for suspect identification [60] and a predictive policing system for the prevention of domestic violence [61]; those not related to law enforcement include a suicide prevention system for veterans [62], a biometric identification system for welfare administration [63], an algorithmic staffing system in a care home for the elderly [64] and a ranking system for patients waiting for a liver transplant [65]. However, it should be noted that the role of AI is not clear in all of these.

In the **hazardous guidance** category, the most straightforward incidents are those where one or more people were killed as a consequence of following navigation directions provided by a geolocation application. The other subcategory of incidents within this category comprises cases where a generative AI system provides the user with unsafe information that could, if acted upon, result in fatalities; however, in most cases this is only presented as a possibility and there is no report of an incident where tangible real-world harm was inflicted. An exception to this is an incident report referring to anecdotal evidence that an AI-generated mushroom foraging guidebook led to a family consuming poisonous mushrooms and being hospitalised [66].

The **algorithmic recommendations** category includes one case where Amazon's algorithm was criticised for recommending products known to have been used in suicide attempts [67], but apart from this, the incidents in this category involve content recommendation algorithms on social media

platforms. Some reports observe the promotion of violence or self-harm on social media without linking it with any specific incidents, but there are also several reports of deaths by accident or suicide where the victims were allegedly influenced by content promoted by TikTok [68, 69] or Instagram [70], as well as two where YouTube is blamed for having contributed to the radicalisation of perpetrators of terrorist attacks [71, 72]. Social media platforms also use AI algorithms for content moderation, the failure of which has been linked with deaths by drug overdose [73] and murder [74].

The incidents in the final category, **conversational agents**, are those where the AI system is acting in a social role, serving as a companion rather than a mere information source. Despite the attempts of developers to build guardrails into their systems, there are several documented cases of them encouraging violence or self-harm in the interactions with the user. For example, [75] describes an incident where a chatbot suggested that a teenage boy kill his parents, and in another incident, a man attempted to assassinate Queen Elizabeth II after being encouraged by a chatbot [76]. There are two known cases where interactions with a chatbot have been suspected of driving a person to suicide [77, 78]. A common feature of the last three incidents is that in each one, the individual in question had reportedly formed an emotional – even romantic in at least one case – attachment to the chatbot.

5. Discussion

Neither the literature review nor the review of AI incidents presented in this paper is intended to be taken as exhaustive. In terms of breadth, incorporating more sources of material and refining the search strategy would help make the big picture more comprehensive and less biased. In terms of depth, there are many topics on which a closer examination of the relevant papers and incident reports would be warranted to gain further insight into the academic and journalistic discourse on AI and death.

Another thing worth noting here is that the set of themes discussed in Section 4 is the product of a process with a substantial inherent subjective element and represents but one way of interpreting and organising the source material. A more rigorous process involving multiple coders would presumably result in a more robust categorisation and allow for more in-depth cross-domain analysis between incident reports and academic literature. Nevertheless, the results of the research presented here lend themselves to a number of interesting observations, along with some specific directions for future research.

The literature review in Section 3, though limited in scope, indicates that research at the intersection of AI, ethics and death covers a diverse range of topics, among which the safety of AI systems is not particularly prominent. Instead, the reviewed literature is dominated by questions such as how AI can help prolong life, what is the capacity of AI systems to make moral decisions involving death and how AI is changing the way we relate to death and the dead. This does not mean that AI safety is being ignored as a research topic, but does suggest that the research is predominantly discussed in such terms that the search fails to find it. This, in turn, could be symptomatic of blind spots in terms of recognising the full implications of the work, but further research is needed to adequately explore this hypothesis.

Comparing the literature with the AI incidents reviewed in Section 4, it can be seen that there is a considerable degree of overlap. The ethically controversial nature of AI-controlled weapon systems is prominently featured in both, as are problems associated with AI systems that make potentially life-or-death decisions in the administration of justice, medical care or other essential services. In other areas, a separation of concerns can be observed; for example, whereas the academic discourse on autonomous vehicles devotes much attention to the idea of such vehicles making value-based choices, incident reports involving them tend to focus on the more concrete issues of safety and liability, which (unlike hypothetical trolley problem scenarios) are immediately relevant to the technology currently in use. Nevertheless, the incident reports do also raise questions of a more philosophical nature regarding the ethics of deploying self-driving technology in the real world.

One notable difference in the discourse between academic literature and incident reports is in the area of social media algorithms. In the former, these barely came up at all, whereas among the latter, several documented incidents involving highly popular platforms were found. Again, this does not

indicate that the safety of these algorithms is being ignored altogether, and it is also worth noting that a handful of incidents is not enough to prove the existence of a causal link between the algorithms and death. There is active debate in the literature on e.g. the effects of social media use on mental health, but this is taking place in such a context that it falls outside the scope of the literature review carried out for this paper.

Another area of divergence is generative AI, more specifically chatbots. Regarding these, in the literature there is much more interest in griefbots, and more generally in various conceptual (rather than causal) connections between chatbots and death. There is some overlap with the incident reports, some of which do bring up ethical controversies associated with the digital resurrection of loved ones or celebrities, but predominantly the reports deal with the (sometimes actualised) risk of tangible harm resulting from chatbots making unsafe suggestions.

As noted above, AI safety is by no means being ignored, and generative AI is no exception. However, as demonstrated by the examples given in the previous section – which represent but a small fraction of all generative AI incidents – the guardrails implemented in existing systems are far from foolproof. Quite apart from the question of how serious the developers of these systems really are about the safety of their products, dialogue safety is a subtle and subjective concept, perceptions of which are affected by various socio-cultural factors, making it challenging to establish the ground truth for guardrail models intended to block unsafe interactions [79]. Moreover, it has been argued that for the purpose of building ethical chatbots, the concept of safety is inadequate to begin with and should be replaced with something more holistic [80].

The two suicide cases associated with chatbots highlight the difficulty of determining what is permissible and desirable behaviour for an AI system designed to interact socially with humans. Incidents such as these suggest that guardrail models should be trained to detect signs of the user getting emotionally invested in the chatbot in an unhealthy way, which is likely to present a major challenge. Assuming that AI companionship has potential benefits and that emotional attachment plays a role in these, it may not be desirable to eliminate it altogether, but to minimise the risks, the phenomenon requires fine calibration. To address this issue, a cross-disciplinary research effort is needed.

6. Conclusion

This paper examined the safety of AI systems through the lens of documented incidents that resulted, or could have resulted, in loss of human life. While AI systems that operate in the physical realm such as self-driving cars are perhaps the most obvious possibility, it was found that AI may contribute to death in a variety of other ways as well: by making decisions that cause denial of life-sustaining services, by providing erroneous guidance on safety-critical topics, by recommending harmful content or by persuading the user to commit harmful acts. Alongside the incident reports, academic literature on AI and death was reviewed; it was observed that the questions addressed in the literature overlap with the incident reports in some areas while diverging in others, conversational agents being one example of the latter.

Self-driving cars and conversational agents are two very different applications of AI technology, but they both highlight a fundamental discontinuity between the external appearance and internal logic of AI systems. Externally, their capabilities are increasingly indistinguishable from – or superior to – those of humans in many situations, creating a temptation for companies to roll out new products and features quickly and for consumers to rely on them extensively. Meanwhile, internally the systems are radically different from the human mind, complicating efforts to understand their limitations and foresee ways in which they might inflict harm. A techno-centric approach to AI safety, focused on perfecting machine learning models without addressing this gap, is therefore likely to fail.

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] R. May, J. Krüger, T. Leich, SoK: How Artificial-Intelligence Incidents Can Jeopardize Safety and Security, in: Proceedings of the 19th International Conference on Availability, Reliability and Security, 2024, pp. 1–12. doi:10.1145/3664476.3664510, article no. 44.

[2] W. Salhab, D. Ameyed, F. Jaafar, H. Mcheick, A Systematic Literature Review on AI Safety: Identifying Trends, Challenges, and Future Directions, IEEE Access 12 (2024) 131762–131784. doi:10.1109/ACCESS.2024.3440647.

[3] B. J. Copeland, artificial intelligence, 2025. URL: https://www.britannica.com/technology/artificial-intelligence, accessed 31 August, 2025.

[4] The Responsible AI Collaborative, Artificial Intelligence Incident Database, n.d. URL: https://incidentdatabase.ai/, accessed 23 April, 2025.

[5] AIAAIC, AIAAIC Repository, n.d. URL: https://www.aiaaic.org/aiaaic-repository, accessed 23 April, 2025.

[6] OECD, OECD AI Incidents and hazards Monitor, n.d. URL: https://oecd.ai/en/incidents, accessed 23 April, 2025.

[7] V. Braun, V. Clarke, Using thematic analysis in psychology, Qualitative Research in Psychology 3 (2006) 77–101. doi:10.1191/1478088706qp063oa.

[8] H. Smith, Artificial Intelligence for Clinical Decision-Making: Gross Negligence Manslaughter and Corporate Manslaughter, The New Bioethics 30 (2024) 228–242. doi:10.1080/20502877.2024.2416862.

[9] R. Vandersluis, J. Savulescu, The selective deployment of AI in healthcare, Bioethics 38 (2024) 391–400. doi:10.1111/bioe.13281.

[10] K. J. Cios, G. William Moore, Uniqueness of medical data mining, Artificial Intelligence in Medicine 26 (2002) 1–24. doi:10.1016/S0933-3657(02)00049-0.

[11] G. Pontes, A. Duarte, D. Cuevas, M. Salazar, M. Miranda, A. Abelha, J. M. Machado, A moral decision support system in medicine, in: 25th European Simulation and Modelling Conference, ESM 2011, 2011.

[12] J. Chen, A. Cadiente, L. J. Kasselman, B. Pilkington, Assessing the performance of ChatGPT in bioethics: a large language model's moral compass in medicine, Journal of Medical Ethics 50 (2024) 97–101. doi:10.1136/jme-2023-109366.

[13] S. Yadav, N. Singh, K. V. R. Devi, G. Nijhawan, R. S. Zabibah, A. K, Synchronizing Innovation and Accountability in the Ethical Consequences of Artificial Intelligence, in: 2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON), volume 10, 2023, pp. 397–402. doi:10.1109/UPCON59197.2023.10434644.

[14] T. A. Mac, Bias and discrimination in ML-based systems of administrative decision-making and support, Computer Law & Security Review 55 (2024) 106070. doi:10.1016/j.clsr.2024.106070.

[15] S. Robbins, Recommending Ourselves to Death: Values in the Age of Algorithms, International Library of Ethics, Law and Technology 40 (2023) 147–161. doi:10.1007/978-3-031-34804-4_8.

[16] L. Eggert, Autonomised harming, Philosophical Studies 182 (2025) 1–24. doi:10.1007/s11098-023-01990-y.

[17] L. T. Bergmann, Ethical Issues in Automated Driving—Opportunities, Dangers, and Obligations, in: A. Riener, M. Jeon, I. Alvarez (Eds.), User Experience Design in the Era of Automated Driving, Springer International Publishing, Cham, 2022, pp. 99–121. doi:10.1007/978-3-030-77726-5_5.

[18] J. Wanner, L.-V. Herm, M. Langer, F. Imgrund, C. Janiesch, A moral consensus mechanism for autonomous driving: Towards a law-compliant basis of logic programming, in: 15th International

Conference on Business Information Systems 2020: Developments, Opportunities and Challenges of Digitization, WIRTSCHAFTSINFORMATIK 2020, 2020. doi:10.30844/wi_2020_a2.

[19] J. Rhim, J.-H. Lee, M. Chen, A. Lim, A Deeper Look at Autonomous Vehicle Ethics: An Integrative Ethical Decision-Making Framework to Explain Moral Pluralism, Frontiers in Robotics and AI 8 (2021). doi:10.3389/frobt.2021.632394.

[20] R. Kopecky, M. Jirout Košová, D. D. Novotný, J. Flegr, D. Černý, How virtue signalling makes us better: moral preferences with respect to autonomous vehicle type choices, AI & Society 38 (2023) 937–946. doi:10.1007/s00146-022-01461-8.

[21] E. P. Bjørgen, S. Madsen, T. S. Bjørknes, F. V. Heimsæter, R. Håvik, M. Linderud, P.-N. Longberg, L. A. Dennis, M. Slavkovik, Cake, Death, and Trolleys: Dilemmas as benchmarks of ethical decision-making, in: Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 2018, pp. 23–29. doi:10.1145/3278721.3278767.

[22] J. A. Oravec, Robots as the Artificial "Other" in the Workplace: Death by Robot and Anti-Robot Backlash., Change Management: An International Journal 21 (2021) 65–78. doi:10.18848/2327-798X/CGP/v21i02/65-78.

[23] E. Riesen, The Moral Case for the Development and Use of Autonomous Weapon Systems, Journal of Military Ethics 21 (2022) 132–150. doi:10.1080/15027570.2022.2124022.

[24] A. Roden-Bow, Killer Robots and Inauthenticity: A Heideggerian Response to the Ethical Challenge Posed by Lethal Autonomous Weapons Systems, Conatus - Journal of Philosophy 8 (2023) 477–486. doi:10.12681/cjp.34864.

[25] B. Akkuş, The Cosmopolitan Principle of Humanity in an Age of Autonomous Weapons: The Role of the Ethical Underpinnings, in: Cyber Security in the Age of Artificial Intelligence and Autonomous Weapons, CRC Press, 2024.

[26] M. Bodziany, Ethical Conditions of the Use of Artificial Intelligence in the Modern Battlefield—Towards the "Modern Culture of Killing", in: A. Visvizi, M. Bodziany (Eds.), Artificial Intelligence and Its Contexts: Security, Business and Governance, Springer International Publishing, Cham, 2021, pp. 45–61. doi:10.1007/978-3-030-88972-2_4.

[27] M. Matsumoto, K. Arai, What's wrong with "Death by Algorithm"? Classifying dignity-based objections to LAWS, AI & Society (2024). doi:10.1007/s00146-024-02093-w.

[28] M. Chiodo, D. Müller, M. Sienknecht, Educating AI developers to prevent harmful path dependency in AI resort-to-force decision making, Australian Journal of International Affairs 78 (2024) 210–219. doi:10.1080/10357718.2024.2327366.

[29] T. de Swarte, O. Boufous, P. Escalle, Artificial intelligence, ethics and human values: the cases of military drones and companion robots, Artificial Life and Robotics 24 (2019) 291–296. doi:10.1007/s10015-019-00525-1.

[30] K. Stowers, K. Leyva, G. M. Hancock, P. A. Hancock, Life or Death by Robot?, Ergonomics in Design 24 (2016) 17–22. doi:10.1177/1064804616635811.

[31] N. F. Lindemann, The Ethics of 'Deathbots', Science and Engineering Ethics 28 (2022) 60. doi:10.1007/s11948-022-00417-x.

[32] T. Hollanek, K. Nowaczyk-Basińska, Griefbots, Deadbots, Postmortem Avatars: on Responsible Applications of Generative AI in the Digital Afterlife Industry, Philosophy & Technology 37 (2024) 63. doi:10.1007/s13347-024-00744-w.

[33] B. Jiménez-Alonso, I. Brescó de Luna, Chapter 9 - AI and grief: a prospective study on the ethical and psychological implications of deathbots, in: S. Caballé, J. Casas-Roma, J. Conesa (Eds.), Ethics in Online AI-based Systems, Intelligent Data-Centric Systems, Academic Press, 2024, pp. 175–191. doi:10.1016/B978-0-443-18851-0.00011-1.

[34] H. Rodríguez Reséndiz, J. Rodríguez Reséndiz, Digital Resurrection: Challenging the Boundary between Life and Death with Artificial Intelligence, Philosophies 9 (2024) 71. doi:10.3390/philosophies9030071.

[35] T. Tietz, F. Pichierri, M. Koutraki, D. Hallinan, F. Boehm, H. Sack, Digital Zombies - the Reanimation of our Digital Selves, in: Companion Proceedings of the The Web Conference 2018, WWW '18, 2018, pp. 1535–1539. doi:10.1145/3184558.3191606.

[36] P. Ulguim, Digital Remains Made Public: Sharing the dead online and our future digital mortuary landscape, AP: Online Journal in Public Archaeology 8 (2018) 153–176. doi:10.23914/ap.v8i2.162.
[37] A. Orita, "Resurrecting" dead celebrities: Editing and using remnant data of the deceased through AI, in: 2022 IEEE International Symposium on Technology and Society (ISTAS), volume 1, 2022, pp. 1–4. doi:10.1109/ISTAS55053.2022.10227098.
[38] H. Lowenstein, N. Steinfeld, H. Rosenberg, The Ethical Implications of Prosocial Synthetic Resuscitation: Analysing User Comments to a Deepfake Campaign Addressing Intimate Partner Violence, Journal of Creative Communications 20 (2025) 23–40. doi:10.1177/09732586241276984.
[39] N. S. Jecker, R. Sparrow, Z. Lederman, A. Ho, Digital Humans to Combat Loneliness and Social Isolation: Ethics Concerns and Policy Recommendations, Hastings Center Report 54 (2024) 7–12. doi:10.1002/hast.1562.
[40] J. Banks, Deletion, departure, death: Experiences of AI companion loss, Journal of Social and Personal Relationships 41 (2024) 3547–3572. doi:10.1177/02654075241269688.
[41] A. L. Guzman, Should Machines Write About Death? Questions of Technology, Humanity, and Ethics in the Automation of Journalism, in: S. J. A. Ward (Ed.), Handbook of Global Media Ethics, Springer International Publishing, Cham, 2021, pp. 555–574. doi:10.1007/978-3-319-32103-5_28.
[42] A. Fitzgerald, Why Synthetic Data Can Never Be Ethical: A Lesson from Media Ethics, Surveillance & Society 22 (2024) 477–482. doi:10.24908/ss.v22i4.18324.
[43] A. Sigurgeirsson, E. L. Ungless, Just Because We Camp, Doesn't Mean We Should: The Ethics of Modelling Queer Voices, 2024. doi:10.48550/arXiv.2406.07504, arXiv:2406.07504 [cs].
[44] T. Lorenc, Artificial intelligence and the ethics of human extinction, Journal of Consciousness Studies 22 (2015) 194–214.
[45] S. Shetty, N. Joshi, A. Banubakode, Shifting from Red AI To Green AI, in: Artificial Intelligence, Machine Learning and User Interface Design, Bentham Science Publishers, 2024, pp. 191–209.
[46] D. Budde, M. Alonso-Villota, S. Pickl, Ethics and the Threat to Infrastructure, in: S. Pickl, B. Hämmerli, P. Mattila, A. Sevillano (Eds.), Critical Information Infrastructures Security, Springer Nature Switzerland, Cham, 2024, pp. 41–61. doi:10.1007/978-3-031-62139-0_3.
[47] OECD, Tesla's Full Self-Driving features undergo probe by NHTSA, n.d.. URL: https://oecd.ai/en/incidents/108306, accessed 23 April, 2025.
[48] OECD, Makers of Self-Driving Cars Ask What to Do With Human Nature, n.d.. URL: https://oecd.ai/en/incidents/1009, accessed 23 April, 2025.
[49] W. Frick, Tesla, Autopilot, and the Challenge of Trusting Machines, Harvard Business Review (2016). URL: https://hbr.org/2016/07/tesla-autopilot-and-the-challenge-of-trusting-machines, accessed 23 April, 2025.
[50] OECD, Autonomous cars vs human driver?, n.d. URL: https://oecd.ai/en/incidents/8437, accessed 23 April, 2025.
[51] M. Hiltzik, Tesla's aggressiveness is endangering people, as well as the cause of good government, 2021. URL: https://www.latimes.com/business/story/2021-10-27/now-tesla-fans-are-endangering-good-government, accessed 23 April, 2025.
[52] R. Mitchell, Tesla touts self-driving to consumers. To the DMV, it tells a different tale, 2021. URL: https://www.latimes.com/business/story/2021-03-09/elon-musk-wants-it-both-ways-with-telsas-full-self-driving, accessed 23 April, 2025.
[53] D. Weitzner, Push for AI innovation can create dangerous products, 2022. URL: https://econotimes.com/Push-for-AI-innovation-can-create-dangerous-products-1637689, accessed 23 April, 2025.
[54] AIAAIC, Kargu-2 fully autonomous drone attacks libyan armed forces, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/kargu-2-autonomous-drone-attack, accessed 23 April, 2025.
[55] AIAAIC, Russian KUB-BLA drones conduct autonomous 'suicide' attacks, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/russian-kub-bla-suicide-drone-attacks, accessed 23 April, 2025.

[56] AIAAIC, Police use of robot to kill Dallas shooting suspect prompts controversy, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/police-robot-kills-dallas-shooting-suspect, accessed 23 April, 2025.
[57] AIAAIC, Israel accused of using AI 'target factory' to slaughter Palestinian women, children, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/israel-uses-habsora-24-hour-automated-target-factory-against-palestinians, accessed 23 April, 2025.
[58] AIAAIC, Israel reportedly uses AI to identify 37,000 Hamas targets, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/israel-reportedly-uses-ai-to-identify-37000-hamas-targets, accessed 23 April, 2025.
[59] AIAAIC, Adam toledo killed by chicago police using shotspotter, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/adam-toledo-killed-by-chicago-police-using-shotspotter, accessed 23 April, 2025.
[60] AIAAIC, Mohammed Khadeer facial recognition wrongful arrest, death, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/mohammed-khadeer-facial-recognition-wrongful-arrest-death, accessed 23 April, 2025.
[61] AIAAIC, Murdered Spanish woman Lobna Hehmid wrongly assessed by domestic gender violence algorithm, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/lobna-hehmid-wrongly-assessed-by-gender-violence-algorithm, accessed 23 April, 2025.
[62] AIAAIC, US Veterans Affairs suicide prevention algorithm favours white men, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/virginia-suicide-prevention-algorithm-favours-white-men, accessed 23 April, 2025.
[63] AIAAIC, Aadhaar glitches result in villagers' starvation, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/aadhaar-glitch-results-in-villagers-starvation, accessed 23 April, 2025.
[64] D. Atherton, Incident 716: Algorithmic Staffing Failures Linked to Resident Deaths at Leading Assisted-Living Chain Brookdale, n.d. URL: https://incidentdatabase.ai/cite/716, accessed 23 April, 2025.
[65] AIAAIC, Native Americans are shut out of us liver transplant system, n.d. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/native-americans-are-shut-out-of-us-liver-transplant-system, accessed 23 April, 2025.
[66] OECD, Ripple CTO highlights AI controversy over dangerous Mushroom identification book | AI | CryptoRank.io, n.d. URL: https://oecd.ai/en/incidents/100070, accessed 23 April, 2025.
[67] AIAAIC, Amazon sales of suicide chemical compound is questioned, n.d. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/amazon-chemical-food-preservative-suicides, accessed 23 April, 2025.
[68] K. Lam, Incident 286: TikTok's "For You" Allegedly Pushed Fatal "Blackout" Challenge Videos to Two Young Girls, n.d. URL: https://incidentdatabase.ai/cite/286, accessed 23 April, 2025.
[69] D. Atherton, Incident 869: TikTok Algorithms Allegedly Linked to Minors' Exposure to Harmful Content, n.d. URL: https://incidentdatabase.ai/cite/869, accessed 23 April, 2025.
[70] K. Lam, Incident 382: Instagram's Exposure of Harmful Content Contributed to Teenage Girl's Suicide, n.d. URL: https://incidentdatabase.ai/cite/382, accessed 23 April, 2025.
[71] S. McGregor, K. Lam, Incident 89: The Christchurch shooter and YouTube's radicalization trap, n.d. URL: https://incidentdatabase.ai/cite/89, accessed 23 April, 2025.
[72] S. McGregor, Incident 476: YouTube Recommendations Allegedly Promoted Radicalizing Material Contributing to Terrorist Acts, n.d. URL: https://incidentdatabase.ai/cite/476, accessed 23 April, 2025.
[73] D. Atherton, Incident 758: Teen's Overdose Reportedly Linked to Meta's AI Systems Failing to Block Ads for Illegal Drugs, n.d. URL: https://incidentdatabase.ai/cite/758, accessed 23 April, 2025.
[74] AIAAIC, Professor Meareg Amare Abrha murder prompts USD 2 billion legal challenge, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/

professor-meareg-amare-abrha-doxxing-murder, accessed 23 April, 2025.
[75] AIAAIC, Character AI chatbot suggests son kills his parents, n.d.. URL: https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/character-ai-chatbot-suggests-son-kills-his-parents, accessed 23 April, 2025.
[76] D. Atherton, Incident 569: Chatbot Encourages Man to Plot Assassination of Queen Elizabeth II, n.d. URL: https://incidentdatabase.ai/cite/569, accessed 23 April, 2025.
[77] S. McGregor, Incident 505: Man Reportedly Committed Suicide Following Conversation with Chai Chatbot, n.d. URL: https://incidentdatabase.ai/cite/505, accessed 23 April, 2025.
[78] D. Atherton, Incident 826: Character.ai Chatbot Allegedly Influenced Teen User Toward Suicide Amid Claims of Missing Guardrails, n.d. URL: https://incidentdatabase.ai/cite/826, accessed 23 April, 2025.
[79] L. Aroyo, A. Taylor, M. Díaz, C. Homan, A. Parrish, G. Serapio-García, V. Prabhakaran, D. Wang, DICES Dataset: Diversity in Conversational AI Evaluation for Safety, Advances in Neural Information Processing Systems 36 (2023) 53330–53342.
[80] H. Kempt, A. Lavie, S. K. Nagel, Towards a Conversational Ethics of Large Language Models, American Philosophical Quarterly 61 (2024) 339–354. doi:10.5406/21521123.61.4.04.