Studi di estetica, anno LIII, IV serie, 2/2025 Sensibilia
ISSN 0585-4733, ISSN digitale 1825-8646, DOI 10.7413/1825864723

Ermelinda Rodilosso

When AI decides
Ethical challenges in life-and-death scenarios

Abstract
This paper aims to shed light on the often-overlooked ethical tensions between Artificial Intelligence (AI) and human decision-making in life-and-death scenarios, where outcomes directly affect human life, survival, and well-being. Focusing on the issues of autonomy and control, I argue that while AI systems can exhibit computational autonomy, they lack the deliberative and moral capacities required for full ethical agency. Through case studies in military and healthcare contexts, I argue for the necessity of consistent human control in ethically sensitive decisions.

Keywords
AI ethics, Applied ethics, Technology

Received: 12/03/2025
Approved: 04/04/2025
Editing by: Ermelinda Rodilosso

ermelinda.rodilosso@outlook.it (Università di Roma Tor Vergata)

1. *Introduction*

Artificial intelligence has reached a level of sophistication that would have been unthinkable just a few decades ago. Its versatility has encouraged its integration into a growing number of industries, from marketing to entertainment, from transportation to creative work. Today, we can process vast volumes of information to obtain statistical data on purchasing habits or political preferences, predict the evolution of weather patterns and fluctuations in rental prices, and generate financial plans, programming codes, and even hyper-realistic images. Moreover, AI can perform these tasks within seconds, significantly reducing the time and effort required for human activities. It is not surprising, then, that artificial intelligence is becoming an integral part of most human experiences. Its ability to mimic – and sometimes even surpass – human intelligence makes our tasks easier, faster, and more efficient.

However, while the benefits of artificial intelligence are evident enough to justify its adoption in almost every area of human life, there is far less awareness of its potential harm. This disparity is largely due to the unprecedented pace of technological advancement, which surpasses our ability to conceive new questions, new methods of inquiry, and, consequently, critical reflections on the consequences of AI development. We are facing levels of sophistication and complexity that are difficult to understand, manage, or predict – especially in a context characterized by constant and dynamic change. For this reason, speculative thinking that investigates the nature and effects of these tools on human and non-human lives is crucial. This kind of inquiry, given its multifaceted and evolving nature, requires a radical rethinking of the traditional questions and categories used in disciplines such as epistemology, philosophy of mind, and cognitive sciences, as well as ethics, political philosophy, and the social sciences. Indeed, if human-machine interactions could once be analyzed through relatively simple categories and frameworks, we must now develop brand-new theoretical and practical tools to manage and solve dilemmas that have relevant ethical significance in human existence.

In this paper, I focus on a relevant issue in AI ethics: the impact that artificial intelligence can have on human decision-making and control. Specifically, I will address the case of autonomous systems in the military and healthcare sectors. These are areas where difficult decisions can determine life or death, making the role of artificial intelligence in supporting such decisions particularly controversial. One might argue that decisions of such gravity should remain strictly within human control, given

the severe consequences involved. Allowing a machine or algorithm to decide matters of life and death is intuitively both immoral and dangerous – a statement that would not require lengthy arguments. However, in this work I will present concrete examples of how AI systems can shape life-and-death scenarios while also claiming the need to impose moral principles that limit the decision-making autonomy of AI. There are situations where artificial intelligence, by default, should never replace human decision-making processes, as doing so risks leading to harmful consequences, including death and the experience of pain. I will outline practical cases in which these principles are underestimated and offer insights into the dangers of an uncritical and overly optimistic reliance on AI in critical decision-making processes.

2. *The ethical significance of AI in decision-making*

Any work that seeks to investigate the intersection between human decision-making and artificial intelligence may raise numerous doubts and questions. Have we truly reached such a high level of innovation to the point that AI can influence our decision-making processes? Can artificial intelligence make decisions for us? Does AI possess a form of agency? If artificial intelligence is advanced enough to shape our decisions or even make them on our behalf, and if it is endowed with a form of agency, could its influence be truly decisive in moral contexts that require complex deliberation? These are all legitimate questions that must be carefully considered, as they determine the legitimacy and foundation of this type of investigation.

As regards the first two questions, I think providing practical examples shows more clearly the co-implications that link human decision-making and artificial intelligence. As I already mentioned, artificial intelligence is massively present in our everyday experiences: we encounter it while looking for information, entertainment, and even social interactions (Ilyas 2022). Algorithms filter the information we see in search engines and social media, recommend films or books based on our preferences, and suggest people we might want to meet and/or start a romantic relationship with. But can we say that artificial intelligence makes actual "decisions" in these cases? The answer depends on the perspective we adopt. If we define decision-making as the process of selecting between multiple options – choosing between element A, element B, or element C – then the answer is yes. Algorithms analyze data and, based on their findings, "decide"

which kind of strategy to follow and which alternative to present to us amidst a huge range of choices. However, this type of decision-making is largely uncritical and does not require a complex deliberative process. Conversely, if we define decision-making as a deliberative process, then the actions performed by AI systems do not constitute true decision-making. They lack several key characteristics that make human decisions complex deliberative processes. As works such as *Upheavals of Thought* by Martha Nussbaum (2001) and *Emotional Intelligence* by Daniel Goleman (1995) show, decision-making involves not only evaluation and calculation but also emotions, contextual experience, and intentionality – elements that AI clearly lacks (Yıldız 2025).

However, we should not assume that, simply because AI-driven processes differ from human decision-making, they have no influence on the choices we actively make. It's quite the opposite. The options presented to us by algorithms have a significant impact on what we decide every day. This influence is particularly evident in areas such as information and marketing. The information we are exposed to in search engines and social media is not random: it is the result of a highly selective process designed to present certain types of content while excluding others because they are deemed more relevant, relatable, or appealing (Zhang et al. 2021). The same applies to marketing. Through profiling processes, machine learning algorithms identify our preferences, predict what we might want to buy, and, based on this analysis, suggest specific products while ignoring others. The mechanisms underlying algorithms that filter information and those that tailor product recommendations operate in a somewhat similar way (Habil *et al.* 2023).

These processes have substantial repercussions on our decision-making. If we are constantly exposed to a particular type of information or product – while alternative perspectives or choices remain obscured – our decisions will have very different outcomes in accordance with what is shown to us. For instance, machine learning algorithms on social media select content in alignment with users' preferences and, in the case of political preferences, can even spur ideological polarization (Rodilosso 2024). While AI does not make these decisions in place of humans, it undoubtedly shapes the experiential and decisional landscape within which we absorb certain information, stimuli, and experiences and, in the light of them, make a certain decision.

The functioning of recommendation algorithms in the fields of information and marketing shows that, at an underlying level, AI is now capable

of influencing our decisions, though reasonable doubts remain about the extent and limits of this capacity. While one can argue about the actual effectiveness and scope of these tools, these systems have some kind of *indirect moral impact* on human beings, as the selections made by algorithms and machines currently shape the experiential ecosystems in which human decision-making takes place.

However, this is only one of the ways in which artificial intelligence and decision-making can intersect. There are circumstances in which algorithms and machines' responses have a *direct moral impact* on human beings. In this instance, AI's decisions are taken without any human control and, at the same time, have immediate, morally significant consequences on human beings. When AI "decides" for us, their calculated choices are not limited to obliquely directing our opinions and preferences, but to replace human agency and control. In this paper, I will focus on the second type of AI impact: cases where artificial intelligence replaces human decision-making in morally critical contexts. I will specifically examine how autonomous systems in the military and healthcare sectors can operate independently of human oversight. Given the scope and immediacy of the risks these technologies pose to human welfare and life, an analysis that crosses normative theories and real-world scenarios is becoming ever more urgent.

3. Computational autonomy vs. human control

When exploring the ethics of artificial intelligence – autonomous systems in particular – there are two key words that we should consider as pivotal: autonomy and control. These concepts are ethically significant because they define how we assign accountability and moral responsibility to machines. An autonomous system that acts without sufficient human control risks creating gaps in responsibility: who is to blame if harm occurs? If an AI system makes a decision that leads to harm or death, is the developer responsible, the owner of the system, or the system itself? The situation becomes even more ethically complex when computational control begins to displace human control. What happens when algorithmic logic overrides human judgment, or when decisions are made in ways that are opaque to human agents?

As Luciano Floridi has observed, artificial intelligence is not only taking up an expanding space in our lives but is also making "increasingly consequential decisions autonomously", raising urgent questions about the implications of "artificial agency" (Floridi 2025). According to Basti and Vitiello (2023) AI systems can exhibit a form of autonomy that enables them to make

ethically relevant decisions. They argue such systems may be considered "autonomous moral agents" if they show the capacity to integrate "ethical constraints" and if they display a "deontic higher order logic". In contrast, Chakraborty and Bhuyan (2024) claim that even if autonomous systems can operate without human oversight, they do not possess autonomy or the power of reasoning in a Kantian sense.

The level of ethical autonomy of artificial intelligence systems is still much debated and there seems to be no general agreement on this issue. Nevertheless, I believe the distinction proposed by James Moor (2006) on artificial agents can further clarify the reflection. Moor divides artificial intelligence into different artificial agents that possess varying degrees of ethical capacity.

(a) *Ethical-Impact Agents*. Systems that affect humans in ethically significant ways, but do not make ethical decisions themselves.
(b) *Implicit Ethical Agents*. Systems that are designed to avoid unethical outcomes but are not capable of truly understanding ethics.
(c) *Explicit ethical agents*. Systems that are capable of reasoning about ethical principles and making decisions based on them. They can simulate ethical deliberation using formal logic or decision frameworks, but they lack consciousness, emotions, or true understanding.
(d) *Full ethical agents*. Systems that have consciousness, intentionality, free will, and moral responsibility required for genuine moral agency.

I consider this distinction fruitful as it establishes specific criteria and capabilities that different levels of artificial agents must exhibit to be considered not only autonomous, but *ethically* autonomous. In light of these criteria, we can say that, although endowed with sophisticated skills, autonomous systems do not fall into the category of "full ethical agents", but rather the category of "explicit ethical agents", since they are capable of understanding moral issues from a formal standpoint, but cannot fully grasp practical reasoning and human emotions.

This important categorization shows us that if we are to judge AI as morally autonomous, we must not rely solely on deontic reasoning capacities, but also on deliberative ones. This means that such systems would need to demonstrate a deliberative ethical capacity that allows them to respond to significant moral situations. However, as I mentioned in the previous section, these instruments lack certain elements that are crucial in deliberative processes. As noted by Martha Nussbaum (2001, 2011), deliberative ethical decision-making requires practical reason, intentionality, embodied experience, a narrative sense of the self, and empathy. For these reasons, I believe that

although AI systems can make computational decisions – based on algorithmic processing and data evaluation – these differ significantly from *deliberative moral decisions*, which involve self-reflection, empathy, and practical reasoning. Deliberative capacity, which is uniquely human, distinguishes explicit ethical agents from full ethical agents, but also AI's moral impact from its ability to perform proper moral decision-making. This taxonomy offers clear tools for evaluating how much autonomy should be granted to artificial intelligence – particularly in morally complex decisions.

Fiorella Battaglia (2025) effectively examines the issue of control in autonomous systems and how the loss of human control can lead to dehumanization. First, the notion of "control" raises the thorny issue of moral responsibility for these instruments. On the one hand, control is the precondition of responsibility: if we do not exercise control over ourselves and our actions, we cannot be held responsible and accountable. On the other hand, control does not seem to be a sufficient element to attribute moral responsibility to an agent. Autonomous systems may bear causal responsibility, but they cannot be held morally accountable for a certain decision. This is consistent with the multifactorial and ambiguous structure of artificial intelligence agency – in which responsibility and accountability are distributed across multiple human, technical, and artificial actors.

Taking up Fischer and Ravizza's (1998) theory, Battaglia acknowledges that AI cannot be considered morally responsible, only causally responsible. This is due to the connection between moral responsibility and the capacity to identify with and own the internal mechanism of an action. In other words, even if AI can respond to reasons for its actions, its actions never fully originate from and belong to AI alone. It follows that decisions of moral standing made by autonomous systems can lead to dehumanization phenomena, for at least two reasons. First, if control is "a necessary condition for the exercise of agency" (Battaglia 2025) and if the loss of agency leads to deprivation of moral standing, then autonomous artificial agents can appropriate human control and, consequently, deny their agency and moral standing. Second, if autonomous systems cannot perform morally characterized choices made through genuine moral understanding, then the process enacted by machines is a priori flawed and risks making dehumanizing decisions.

Both the issue of autonomy and control demonstrate that autonomous systems, although they can be designed and trained to follow ethical principles and exhibit some degree of autonomy, control, and agency, are not proficient in making decisions that carry significant moral consequences – or at least not without adequate human supervision. Given this radical difference, I argue that there should always be human control in cases where AI-driven

decisions have life-altering consequences for human beings. The following case studies will illustrate how the unsupervised decisions of autonomous systems can impact human welfare and life. By examining real-world examples from domains such as healthcare and military operations, we can better understand the ethical risks posed by delegating morally significant choices to systems that lack deliberative capacity, empathy, and proper accountability.

4. *"The Gospel", "Lavender", "Iron Dome", and the drones*

The first case I examine in this paper concerns the military sector. The potential for AI to influence military decision-making is not new and has been discussed since the early development of these tools. However, recent advancements in AI-powered defense systems have raised new concerns about their direct impact on human life. Today, numerous military strategies massively rely on artificial intelligence, often with little regard for the ethical boundaries that should limit its use when human survival is at stake. Among the most recognizable examples we can find the AI-driven technologies "The Gospel" and "Lavender", employed by the Israeli Defense Forces (IDF) in real-life military operations. Here I would like to focus on their capacity to actively influence – if not replace – human decisions. For this reason, I will refer to them as AI Military Decision Support Systems: AI-driven systems programmed to process vast amounts of data efficiently and instantaneously to identify, target, and, in some cases, execute military strikes with little to no human intervention (Sharma 2024, Gusterson 2024). Given their precision and speed – exponentially greater than that of human operators – they can be considered a true "mass assassination factory" (Abraham 2023), disproportionately amplifying the destructive potential of the already sophisticated military technology at the IDF's disposal.

According to Aviv Kochavi, who has served as Chief of General Staff of IDF, *The Gospel* enabled the identification of 100 targets per day, whereas exclusive human analysis identified only 50 targets over an entire year (Davies *et al.* 2023). These figures highlight how AI, when used for military purposes without adequate human oversight, severely undermines the moral legitimacy of its use. The so-called "targets" are not merely vehicles or buildings; they are often people whose safety can be undermined by machines. These machines, as I have exposed in paragraph 3, are not able to

perform deliberative moral decisions, as they follow data analysis, pre-programmed designs and rules, and learned patterns. Proponents of these AI-driven military strategies emphasize their advantages in terms of speed and efficiency, which allow them to anticipate and outperform opponents (Brumfiel 2023). However, the moral and human cost of this efficiency cannot be ignored. Allowing algorithms to make decisions in warfare without sufficient human intervention – intervention that both slows down and critically monitors these operations – exposes countless human lives to fatal mistakes and raises ethical concerns about the growing capacity for large-scale annihilation. The *Lavender* system, for instance, is reported to have an error rate of 10% (Al Jazeera 2024) – an alarming high percentage given the stakes. Yet, despite this margin of error, it continues to be deployed for high-risk targeting operations.

The destructive potential enabled by AI-driven warfare is something completely new that our minds still struggle to process. AI-enhanced drones deployed by Israel, working in conjunction with analytical systems such as *The Gospel* and *Lavender*, can acquire target information within moments and be used in various forms of attack, including "dropping grenades, firing missiles, conducting suicide missions, and crashing into civilian infrastructure" (Dana 2024). In addition, these AI-powered systems have been employed in devious strategies, such as broadcasting the cries of women and children to lure civilians from their homes, exposing them to targeted attacks (Euro-Mediterranean Human Rights Monitor 2024). These examples demonstrate how the technological advancement of AI in military applications not only increases the efficiency of military operations but also enables tactics that more closely resemble war crimes than lawful combat – if we can even apply the term "lawful" to acts of killing.

At this point, I would like to mention some of the most pressing ethical risks associated with AI in warfare and examine how these risks manifest in the military field. In particular, I will focus on the issues of autonomy, transparency, and accountability, as well as the potential for misuse. Regarding autonomy, so-called "Lethal Autonomous Weapons" (LAWs) are already operational and can be *defensive*, such as the *Iron Dome* missile defense system, which intercepts and destroys incoming rockets, artillery, and mortars (hence classified as counter-RAM or C-RAM) (Johansson and Falkman 2011, Slesinger 2022), or *offensive*, such as the AI-guided drones discussed earlier. While these systems require some minimal de-

gree of human intervention, it remains unclear how much control humans actually have on these tools. There are three recognized levels of human involvement in autonomous weapon systems:

(a) *Human-in-the-loop*, where a human must initiate the machine's action;
(b) *Human-on-the-loop*, where a human can override or abort a machine-initiated action;
(c) *Human-out-of-the-loop*, where no human intervention occurs, and the system operates independently based on its programming and data.

To date, there is no available evidence confirming the deployment of fully lethal autonomous weapons (LAWs). Existing regulations require LAWs to be designed in such a way as to necessitate human judgment. For example, the U.S. Department of Defense's Directive 3000.09, originally issued in 2012 and renewed in 2023, mandates that autonomous and semi-autonomous weapon systems "allow commanders and operators to exercise appropriate levels of human judgment over the use of force" (U.S. Department of Defense 2012: 2). However, this directive is both outdated and ambiguous. It fails to define what constitutes an "appropriate level of human judgment" or how this judgment should be exercised in life-and-death scenarios.

The risk of these systems evolving into autonomous decision-makers operating with minimal or no human oversight is increasingly plausible. If AI-driven systems have the potential to make life-and-death decisions without human intervention, they disrupt the moral accountability of military operations, making it difficult to determine who is responsible for specific actions and their consequences. Amanda Sharkey warns that this shift toward automation in warfare risks eroding human dignity by eliminating "the human reflection that is essential for justice, morality, and law" (2019). Excessive machine autonomy could initiate processes that ultimately undermine human autonomy itself, exploring domains that should remain exclusively human, such as the right to life and personal security.

Another major concern is the lack of transparency in AI-driven military systems. The complexity of AI algorithms used for targeting and strikes – combined with the secrecy surrounding military technology – makes it nearly impossible to fully understand how specific decisions are made. This lack of transparency presents significant obstacles to accountability.

Military operations are already difficult to reconstruct due to the chaotic nature of warfare, even when human-controlled weapons are used. The challenge of determining responsibility becomes even greater when it is unclear who initiated an operation or whether the machine acted independently. This creates harmful ambiguities in the domain of moral responsibility. If an autonomous or semi-autonomous LAW commits a war crime or engages in an illicit military operation, who bears responsibility? Should accountability fall on the military personnel operating the system? The politicians who authorized the deployment of AI in warfare? The software engineers who designed the algorithms? Or does responsibility rest with the artificial agency itself? These uncertainties show the highly problematic nature of AI-driven military decision-making and the urgent need for human oversight. Human intervention is crucial for at least three reasons:

(a) *Lack of humanness*. Machines and algorithms lack essential human qualities such as empathy, compassion, remorse (Sancar 2024), and contextual understanding (Christie et al. 2024), all of which are crucial for ethical decision-making. While elaborating a moral judgement, the human mind takes into account context, moral responsibility, and emotions – elements that are alien to artificial intelligence systems.
(b) *Ambiguities in accountability*. AI systems and LAWs present "a challenge for responsibility and accountability" (*Ivi*), as they can replace humans at every stage of a military operation. As I already said, this could make it difficult – if not impossible – to determine who is responsible for war crimes or could lead to the justification of such crimes as the result of non-human errors. This ambiguity creates a dangerous loophole, potentially allowing military personnel and political leaders to violate international law and human rights without consequence.
(c) *Risk of escalation and loss of control*. The speed and efficiency that make artificial intelligence appealing in military contexts also heighten the risk of escalation and instability (Horowitz 2021). Since commands can be executed instantaneously and with extreme precision, any escalation could result in the loss of vast numbers of innocent lives, if not outright massacres.

In light of these concerns, it is evident that AI systems supporting or replacing human decision-making carry significant moral implications and

can exacerbate ethical dilemmas that are far from new. These implications are deeply connected to plausible dehumanization determined by computational autonomy. AI tools can be programmed to execute operations that, while offering a strategic military advantage, ultimately automate decisions that should be carefully weighed and evaluated by human's deliberative capacity and power of reasoning. While these technologies represent a prodigious advancement, allowing algorithms to instantly identify targets and issue commands with little or no human oversight sets a dangerous precedent that deepens the oppression of already vulnerable nations. A prime example of this, again, is the apartheid State of Israel, which, by deploying untested AI systems in warfare, not only strengthens its dominance but also acquires valuable technological resources to tighten its grip on Palestine, escalating what Dana (2024) describes as "genocidal brutality." The use of AI in this context is not merely a matter of military strategy but also a means of pursuing political and ideological objectives.

Another critical issue worth discussing when dealing with artificial intelligence is the role of the digital divide in contexts marked by systemic injustice. The military application of AI exposes marginalized individuals and nations to even more violent forms of domination and oppression. As technological tools become so advanced that they dictate the outcomes of conflicts, it becomes clear that countries lacking the economic resources to develop AI-driven military technologies will have little chance of success. The race to develop cutting-edge AI systems among industrialized nations is widening the gap between wealthy and oppressed countries, reinforcing colonial power structures and exacerbating global inequalities. Artificial intelligence, perhaps more starkly than previous technological innovations, demonstrates how technology can, in some cases, deepen existing disparities rather than serve as a tool for collective liberation. The prospect that these systems will become even more powerful in the future – and that their maintenance will demand immense economic and material resources – suggests that the wars of today and tomorrow will offer no path to justice for oppressed peoples. Instead, AI is increasingly emerging as a vehicle for perpetuating imperialist and colonial ambitions, cementing structural inequalities rather than dismantling them.

5. *Denial, delay, and neglect*

A second, equally problematic case that threatens to undermine human safety in life-and-death decision-making contexts is healthcare. The intersection of artificial intelligence and medical care has proven fruitful in many circumstances. AI is widely used in imaging analysis, where it is trained to interpret medical images and X-rays to identify potential diseases or conditions (Khalifa and Albadawy 2024), improving both the speed and accuracy of diagnoses. It also plays a role in developing personalized treatment plans (Yogeshappa 2024), tailoring medical interventions to individual patients based on their specific needs. Additionally, AI is increasingly used in pharmaceutical research, supporting the discovery of new drugs and active ingredients (Bhattamisra *et al.* 2023). The effectiveness of these tools makes them not only preferable but also indispensable for delivering precise and rapid results that strongly affect people's health and lives. The ability to process and analyze vast amounts of data, detect patterns linking different symptoms and conditions, and assist physicians in formulating accurate diagnoses and treatments demonstrates AI's potential to revolutionize healthcare.

However, in certain instances, the use of artificial intelligence in healthcare does not constitute a promising support system for human decision-making but instead a form of ethical malpractice. AI tools, which should merely enhance the capabilities of healthcare professionals, are being designed to make autonomous or semi-autonomous decisions that directly affect patient care. A particularly notorious case, largely due to its media resonance following the murder of CEO Brian Thompson, is the role of UnitedHealthcare and its misuse of AI in medical decision-making. UnitedHealthcare, currently the most thriving healthcare company in America (Ali and Dobbs 2025), employs advanced data analysis systems to support medical decisions. However, the company has faced important criticism around the way it uses AI to make life-and-death decisions for its patients. One major concern regards the alleged use of the nH Predict System, a software program deputed to analyze patients' clinical data – including diagnoses, conditions, age, and gender. The algorithm cross-references this information with a vast database to identify patterns, generate predictive analyses, and estimate the type of care patients should receive, along with the associated costs (Talia 2024).

According to the American Medical Association, three out of five physicians believe that AI is increasing the number of authorization denials

(American Medical Association 2024) – a concern corroborated by an investigation led by the U.S. Senate Committee. The Committee's findings reveal a correlation between AI-driven decision-making and rising denial rates. Specifically, in UnitedHealthcare's case, denial rates for prior authorization of post-acute care services skyrocketed from 10.9% in 2020 to 16.3% in 2021 and then to 22.7% in 2022 – a period during which the company implemented multiple initiatives to automate the decision-making process. Meanwhile, in Humana's case – another relevant American healthcare company – denial rates were 16 times higher than those of its competitors (U.S. Senate Committee on Homeland Security and Governmental Affairs 2024). These statistics reveal the intent to refine decision-support systems not only to determine which treatments should be denied funding but also to predict which denials were likely to be appealed and which appeals were likely to be overturned (*ivi*). This increasing reliance on automation suggests a deliberate strategy to cut assessment time and related costs, prioritizing efficiency over patient welfare.

What emerges with clarity is the pursuit of financial efficiency at the expense of real patient well-being. The pursuit of efficiency in AI use introduces ethical tensions comparable to those observed in autonomous military systems, particularly regarding accountability and control. In both contexts, life-and-death decisions are being entrusted to algorithms, with human oversight minimized – sometimes to the point of near-total removal. Systems like nH Predict, intended to enhance the quality of medical care, are instead being weaponized to deny care as often as possible for the sake of profit. Here, the same fundamental questions arise as in military AI. How much autonomy should software or machines be granted in making medical decisions? Can human decision-making be entirely replaced by human-out-of-the-loop systems? Does AI autonomy undermine human autonomy and the right to life? The last question is particularly crucial. Denying life-saving or essential medical treatment is, as a matter of fact, a denial of the fundamental human right to life (United Nations General Assembly 1948). It precludes individuals from exercising autonomy and self-determination over their own bodies and well-being. It is evident that decision-making processes that jeopardize human life or degrade its dignity cannot be included autonomous systems' designs. In this area, computational mechanisms demonstrate an obvious incompetence in the field of moral decision-making, which risks impacting on patients' safety.

In light of this, human control in life-and-death medical decisions is essential. The development of human-out-of-the-loop AI systems should

never be permitted in healthcare scenarios because, as in military applications, such systems are prone to error, susceptible to cultural and social biases, incapable of empathy or compassion, and vulnerable to illicit programming and misuse (Zhang and Zhang 2023).

Just as in military AI systems, the issues of transparency and accountability are inextricably linked. AI source codes are proprietary and heavily protected, making the inner workings of these systems opaque and moral accountability ambiguous. When medical treatment is denied, who bears responsibility? The AI developers? The healthcare providers? The insurance companies? The lack of clear accountability not only prevents access to care for those in need but also encourages insurance agencies to perpetuate unjust policies without fear of legal repercussions. The widespread implementation of AI-driven decision systems in healthcare risks reinforcing pre-existing social injustices by perpetuating biases that unjustly harm marginalized communities (Haider *et al.* 2024). Artificial intelligence could further deepen the divide between those who can afford quality healthcare and those who are systematically denied it due to their economic and social status.

All these ethical concerns emphasize the critical importance of human control in life-and-death situations. More broadly, they highlight the necessity of placing human well-being at the center of technological advancement. AI-driven systems that diminish human welfare must be reviewed and redesigned to create meaningful social change instead of being directed to make profits. No technological development can be considered authentic progress if it results in the degradation or dehumanization of human life rather than its improvement. This principle should serve as a foundational standard for any ethical assessment of technology's role in society, particularly given that technology is never entirely neutral in its application. As John Dewey (1925) emphasized, technological development is inherently value-neutral, but it is shaped by human purposes, habits, and societal priorities and worldviews. If current applications of AI in healthcare prioritize profit over patient well-being, efficiency over ethics, and automation over human dignity, then there is an urgent need to change course – redirecting AI development toward decision-making processes that are more humane, just, and respectful of the fundamental rights to life and autonomy.

6. *Ethical risks: What can we do about it?*

So far, this work has provided a brief overview of the ethical risks associated with using AI systems to support human life-and-death decisions. Now, as we approach the conclusion of this analysis, I intend to suggest some possible strategies for mitigating these risks. The first essential step is to avoid *overreliance* on AI and to acknowledge both its fallibility and the human risks involved. The primary reason for the excessive delegation of human decision-making to artificial intelligence lies in overconfidence – a mistaken belief that these systems are infallible. As we have seen, this is far from true: AI-driven decision-making can reflect the same cultural biases and discriminatory tendencies as its developers, it can produce errors and misjudgments, and it lacks the proper skills to perform a proper ethical deliberation that is necessary to make ethical and informed decisions.

Once we recognize the fallibility of AI systems and develop a reasonable skepticism about their trustworthiness in moral decision-making, military, and healthcare agencies must establish robust evaluation and testing mechanisms to ensure that AI's efficiency does not lead to injustice or harm. Third, AI models and source codes should be more transparent and accessible so that the different ways in which the algorithms operate are clearer. This would allow for greater accountability when these systems make mistakes or generate harmful outcomes that affect human welfare.

For ethical risk management to be effective, a comprehensive system of regulatory policies must be developed to define how AI should be employed in military and healthcare settings and, more importantly, what level of involvement AI systems should have in sensitive life-and-death scenarios. Legislation, for the time being, appears to be insufficient and opaque: it requires radical updating to reflect the realities of modern AI advancements rather than remain anchored to outdated conceptions of its potential. Without up-to-date legislation, we not only risk failing to address the ethical concerns outlined here, but we may also find ourselves unprepared for AI misuse scenarios, lacking the tools to respond adequately when they arise. What I believe is most crucial to demand from AI designs is making the role of human intervention mandatory in circumstances where human life is at stake. For this reason, I propose the adoption of the "Meaningful Human Control" (MHC) framework (Santoni de Sio and Van den Hoven 2018) in life-and-death decisions, which would mandate human involvement in any algorithmic operation affecting people's safety or health. MHC is grounded in two key conditions: *tracking* and

tracing. First, autonomous systems must be able to track human moral reasoning – that is, their decisions should be responsive to values that human agents deem ethically relevant in a given context. Second, there must be a traceable link to a human agent who understands the functioning of the system and can be held accountable for its outcomes. In other words, MHC ensures that decisions with important moral consequences are never completely handed over to machines but remain embedded within human understanding and responsibility.

This approach stands in direct contrast to current trends in automation that risk removing humans from decision-making processes and thus dehumanize such processes. Moral reasoning, conscience, the ability to feel remorse, empathy, and a sense of responsibility are fundamental aspects of human decision-making that no machine can replicate. In situations requiring the most excellent exercise of empathy and humanity, we cannot entrust those decisions to computational logic. MHC provides a normative framework that preserves moral responsibility by ensuring that ethically significant decisions remain meaningfully under human control.

7. *Conclusion*

The analysis conducted here was applied-based and aimed to provide insights into the relationship between artificial intelligence and human decision-making in life-and-death situations. The fundamental observation that prompted these reflections is that while we have a growing awareness of AI's benefits, we still lack sufficient ethical tools to fully grasp its scope and potential consequences for human life.

Although AI systems increasingly operate with a high degree of computational autonomy, they remain incapable of moral deliberation in the human sense. As argued in paragraph 3, these systems may process data, simulate ethical reasoning, and even generate decisions with morally significant consequences – yet they do so without the capacities for intentionality, empathy, or self-reflective judgment. These are essential elements of deliberative moral agency and are hitherto unique to human beings. Even if I believe AI is an exceptional tool that should be properly employed for collective human advancement and emancipation – and thus should neither be discouraged nor demonized – we should not give them the status of full ethical moral agents.

The distinctions between moral impact and the capacity of moral decision-making, computational decision and deliberative moral decisions, causal

responsibility and moral responsibility, they all matter deeply when focusing on understanding the relationship between computational autonomy and human control. Without sufficient human control, AI decisions risk generating gaps in responsibility and the possibility of leading to the dehumanization of moral decisions. As such, allowing algorithmic/computational logic to override or displace human decision-making in morally sensitive domains undermines the foundations of ethical responsibility and legal accountability.

For these reasons, I contend that while AI should continue to be developed as a powerful tool for human flourishing, its deployment in morally charged contexts must always be bounded by effective human oversight. Systems that lack the capacity for genuine moral understanding should not be given authority over decisions that bear on human safety, dignity, or life itself. Through this reflection, I have aimed to emphasize the problematic aspects of AI's role in human decision-making, while claiming that the ethical design and governance of AI must include a commitment to preserving human control and deliberative decision-making in life-and-death scenarios. One principle remains non-negotiable: wherever AI operates in ethically significant domains, human responsibility must not be delegated but preserved.

Bibliography

Abraham, Y. *'A mass assassination factory': Inside Israel's calculated bombing of Gaza*, "+972 Magazine", November 30, 2023. Available at: https://www.972mag.com/mass-assassination-factory-israel-calculated-bombing-gaza/ (accessed: 26 February 2024).

Ågerfalk, P.J., *Artificial intelligence as digital agency*, "European Journal of Information Systems", n. 29/1 (2023), pp. 1-8, https://doi.org/10.1080/0960085X.2020.1721947 (accessed: 6 August 2025).

Al Jazeera, *'AI-assisted genocide': Israel reportedly used database for Gaza kill lists*, "Al Jazeera", April 4, 2024. Available at: https://www.aljazeera.com/news/2024/4/4/ai-assisted-genocide-israel-reportedly-used-database-for-gaza-kill-lists (accessed: 26 February 2024).

Ali, S., Dobbs, T., *The promise and peril of AI in healthcare: A cautionary tale of UnitedHealthcare*, "The Bulletin of the Royal College of Surgeons of England", n. 107/2 (2025), pp. 50-105, https://doi.org/10.1308/rcsbull.2025.44 (accessed: 6 August 2025).

American Medical Association, *Physicians Concerned AI Increases Prior Authorization Denials*, "American Medical Association", March 7, 2024. Available at: https://www.ama-assn.org/press-center/press-releases/physicians-concerned-ai-increases-prior-authorization-denials (accessed: 26 February 2024).

Bandura, A., *Toward a psychology of human agency,* "Perspectives on Psychological Science", n. 1/2 (2006), pp. 164-80, https://doi.org/10.1111/j.1745-6916.2006.00011.x (accessed: 6 August 2025).

Basti, G., Vitiello, G., *Deep learning opacity, and the ethical accountability of AI systems. A new perspective*, in eds. R. Giovagnoli, R. Lowe, *The logic of social practices II. Studies in applied philosophy, epistemology and rational ethics*, Springer, 2023.

Battaglia, F., *Dehumanization: An updated philosophical account in the human-machine context*, in eds. J. Seibt, P. Fazekas, O.S. Quick, *Social robots with AI: prospects, risks, and responsible methods*, IOS Press, 2025.

Bhattamisra S.K., Banerjee, P., Gupta, P., Mayuren, J., Susmita, P., Candasamy, M., *Artificial intelligence in pharmaceutical and healthcare research*, "Big Data and Cognitive Computing", n. 7/1 (2023). https://doi.org/10.3390/bdcc7010010 (accessed: 6 August 2025).

Brumfiel, G., *Israel is using an AI system to find targets in Gaza. Experts say it's just the start*, "NPR", December 14, 2023. Available at: https://www.npr.org/2023/12/14/1218643254 (accessed: 26 February 2024).

Bostrom, N., *Superintelligence: Paths, dangers, strategies*, Oxford, Oxford University Press, 2014.

Chakraborty, A., Bhuyan, N., *Can artificial intelligence be a Kantian moral agent? On moral autonomy of AI system*, "AI Ethics", n. 4 (2024), pp. 325-31, https://doi.org/10.1007/s43681-023-00269-6 (accessed: 6 August 2025).

Christie, E.H., Ertan, A., Adomaitis, L. *et al.*, *Regulating lethal autonomous weapon systems: Exploring the challenges of explainability and traceability*, "AI Ethics", n. 4 (2024), pp. 229-45, https://doi.org/10.1007/s43681-023-00261-0 (accessed: 6 August 2025).

Dana, T., *Death dealers: Dynamics of Israel's permanent war economy*, "Capital & Class", 2024, https://doi.org/10.1177/03098168241291350 (accessed: 6 August 2025).

Davies, H., McKernan, B., Sabbagh, D., *'The Gospel': How Israel uses AI to select bombing targets in Gaza*, "The Guardian", December 1, 2023. Available at: https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets (accessed: 26 February 2024).

Dewey, J., *Experience and nature*, New York, Dover Publications, 1925.

Euro-Mediterranean Human Rights Monitor, *Israeli army broadcasts intimidating sounds to lure, kill, and forcibly displace civilians in the Nuseirat camp*, "Euro-Mediterranean Human Rights Monitor", April 16, 2024. Available at: https://euromed-monitor.org/en/article/6271/Israeli-army-broadcasts-intimidating-sounds-to-lure,-kill,-and-forcibly-displace-civilians-in-the-Nuseirat-camp (accessed: 26 February 2024).

Fereidooni, S., Heidt, V. *The fallacy of precision: Deconstructing the narrative supporting AI-enhanced military weaponry*, "Harms and Risks of AI in the Military", https://doi.org/10.18356/6fce2bae-en (accessed: 6 August 2025).

Fischer J.M., Ravizza M., *Responsibility and control: a theory of moral responsibility*. Cambridge University Press, 1998.

Floridi, L., *AI as agency without intelligence: On ChatGPT, large language models, and other generative models,* "Philosophy & Technology", n. 36/1 (2023a), https://doi.org/10.1007/s13347-023-00621-y (accessed: 6 August 2025).

Floridi, L., *The ethics of artificial intelligence: Principles, challenges, and opportunities,* Oxford, Oxford University Press, 2023b.

Floridi, L., *AI as agency without intelligence: On artificial intelligence as a new form of artificial agency and the multiple realisability of agency thesis*, "Philosophy & Technology", n. 38/30 (2025), https://doi.org/10.1007/s13347-025-00858-9 (accessed: 6 August 2025).

Goleman, D., *Emotional intelligence*, New York, Bantam Books, 1995.

Gusterson, H., *It's all Lavender in Gaza*, "Anthropology Today", n. 40/6 (2024), https://doi.org/10.1111/1467-8322.12923 (accessed: 6 August 2025).

Habil, S., El-Deeb, S., El-Bassiouny, N., *AI-based recommendation systems: The ultimate solution for market prediction and targeting*, in ed. C.L. Wang, *The palgrave handbook of interactive marketing*, Palgrave Macmillan, 2023.

Haider, S.A., Borna, S., Gomez-Cabello, C.A. *et al.*, *The algorithmic divide: A systematic review on AI-driven racial disparities in healthcare*, "Journal of Racial and Ethnic Health Disparities", 2024, https://doi.org/10.1007/s40615-024-02237-0 (accessed: 6 August 2025).

Horowitz, M.C., *When speed kills: Lethal autonomous weapon systems, deterrence and stability*, in eds. T.S. Sechser, N. Narang, C. Talmadge, *Emerging technologies and international stability*, London, Routledge, 2022.

Ilyas, M., *Emerging role of artificial intelligence*, "Journal of Systemics, Cybernetics and Informatics", n. 20/6 (2022), pp. 58-65, https://doi.org/10.54808/JSCI.20.06.58 (accessed: 6 August 2025).

Johansson, F., Falkman, G., *Real-time allocation of firing units to hostile targets*, "Journal of Advances in Information Fusion", n. 6/2 (2011), pp. 187-99.

Khalifa, M., Albadawy, M., *AI in diagnostic imaging: Revolutionising accuracy and efficiency*, "Computer Methods and Programs in Biomedicine Update", n. 5 (2024), https://doi.org/10.1016/j.cmpbup.2024.100146 (accessed: 6 August 2025).

Moor, J.H., *The nature, importance, and difficulty of machine ethics*, "IEEE Intelligent Systems", vol. 21, n. 4 (2006), pp. 18-21, https://doi.org/10.1109/MIS.2006.80 (accessed: 6 August 2025).

Newell, A., Simon H.A., *Computer science as empirical inquiry: Symbols and search*, "Communications of the ACM", n. 19/3 (1976), pp. 113-26.

Nussbaum, M., *Creating capabilities: The human development approach*, Harvard University Press, 2011.

Nussbaum, M., *Upheavals of thought*, Cambridge, Cambridge University Press, 2001.

Rodilosso, E., *Filter bubbles and the unfeeling: How AI for social media can foster extremism and polarization*, "Philosophy & Technology", n. 37/71 (2024), https://doi.org/10.1007/s13347-024-00758-4 (accessed: 6 August 2025).

Sancar, I.V., *How can we design autonomous weapon systems?*, "AI Ethics", 2024, https://doi.org/10.1007/s43681-024-00428-3 (accessed: 6 August 2025).

Santoni de Sio, F., van den Hoven, J., *Meaningful human control over autonomous systems: A philosophical account*, "Frontiers Robotics and AI", 5/15 (2018), https://doi.org/10.3389/frobt.2018.00015 (accessed: 6 August 2025).

Sharkey, A., *Autonomous weapons systems, killer robots and human dignity*, "Ethics and Information Technology", n. 21 (2019), pp. 75-87, https://doi.org/10.1007/s10676-018-9494-0 (accessed: 6 August 2025).

Sharma, R.K., *Human-in-the-loop dilemmas The Lavender System in Israel Defence Forces operations*, "Journal of Defence Studies", n. 18/2 (2024), pp. 166-75.

Slesinger, I., *A strange sky: Security atmospheres and the technological management of geopolitical conflict in the case of Israel's Iron Dome*, "The Geographical Journal", n. 188/3 (2022), pp. 429-43, https://doi.org/10.1111/geoj.12444 (accessed: 6 August 2025).

Talia, D., *Algorithms that can deny care, and a call for AI explainability*, "IEEE", n. 57/7 (2024), pp. 109-12, https://doi.org/10.1109/MC.2024.3387012 (accessed: 6 August 2025).

United Nations General Assembly, *The universal declaration of human rights (UDHR)*, "United Nations General Assembly", December 10, 1948. Available at: https://www.un.org/en/about-us/universal-declaration-of-human-rights (accessed: 26 February 2024).

US Department of Defence, *Directive 3000.09. Autonomy in Weapon Systems*, "US Department of Defense", November 21, 2012. Available at: https://web.archive.org/web/20121201105512/http://www.dtic.mil/whs/directives/corres/pdf/300009p.pdf (accessed: 26 February 2024).

U.S. Senate Committee on Homeland Security and Governmental Affairs, *Refusal of recovery: How Medicare advantage insurers have denied patients access to post-acute care*, October 17, 2024. Available at: https://www.hsgac.senate.gov/wp-content/uploads/2024.10.17-PSI-Majority-Staff-Report-on-Medicare-Advantage.pdf (accessed: 26 February 2024).

Yıldız, T., *The minds we make: A philosophical inquiry into theory of mind and artificial intelligence*, "Integrative Psychological and Behavioral Science", n. 59/10 (2025), https://doi.org/10.1007/s12124-024-09876-2 (accessed: 6 August 2025).

Yogeshappa, V.G., *AI-driven precision medicine: Revolutionizing personalized treatment plans*, "International Journal of Computer Engineering & Technology", n. 15/5 (2024), pp. 455-74, https://doi.org/10.5281/zenodo.13841278 (accessed: 6 August 2025).

Zhang, J., Zhang, Z.M., *Ethics and governance of trustworthy medical artificial intelligence*, "BMC Medical Informatics and Decision Making", n. 23/7 (2023), https://doi.org/10.1186/s12911-023-02103-9 (accessed: 6 August 2025).

Zhang, Q., Lu, J. & Jin, Y., *Artificial intelligence in recommender systems*, "Complex & Intelligent Systems", n. 7, pp. 439-57 (2021), https://doi.org/10.1007/s40747-020-00212-w (accessed: 6 August 2025).