Acta Bioethica 2026; 32(1): 41-57. DOI: http://dx.doi.org/10.4067/s1726-569x2026000100041

ARTIFICIAL INTELLIGENCE IN DEFENSE: A BIOETHICAL EXAMINATION OF SOCIETAL RISKS AND GOVERNANCE

Adel Ayed Alshammari[1]

Abstract: This paper explores the societal risks that may arise from the integration of artificial intelligence (AI) into defense innovations and examines how the principle of responsibility can be achieved in this context. Using a qualitative descriptive–analytical approach grounded in international reports and prior studies, the paper identifies four critical domains of societal risk: threats to human life, terrorist misuse, cyber and electronic warfare, and the lack of accountability for actions committed by autonomous or semi-autonomous systems. To address these challenges, a tri-level framework of responsibility is proposed —international, institutional, and socio-cultural— linking global ethical norms with local governance practices and moral values. The study argues that responsible innovation in AI-driven defense technologies requires ethical governance that ensures accountability, human oversight, and societal legitimacy, thereby balancing security imperatives with public trust.

Keywords: Artificial Intelligence, defense, societal risks, governance

Inteligencia Artificial en defensa: un examen bioético de los riesgos sociales y la gobernanza

Resumen: Este artículo explora los riesgos sociales que pueden surgir de la integración de la inteligencia artificial (IA) en innovaciones en defensa y examina cómo se puede lograr el principio de responsabilidad en este contexto. Utilizando un enfoque descriptivo-analítico cualitativo, basado en informes internacionales y estudios previos, el artículo identifica cuatro dominios críticos de riesgo social: amenazas a la vida humana, uso indebido de terroristas, guerra cibernética y electrónica, y la falta de rendición de cuentas por acciones cometidas por sistemas autónomos o semiautónomos. Para abordar estos desafíos, se propone un marco de responsabilidad a tres niveles —internacional, institucional y sociocultural— que vincule las normas éticas globales con las prácticas de gobernanza local y los valores morales. El estudio sostiene que la innovación responsable en tecnologías de defensa impulsadas por IA requiere una gobernanza ética que garantice rendición de cuentas, supervisión humana y legitimidad social, equilibrando así los imperativos de seguridad con la confianza pública.

Palabras clave: Inteligencia Artificial, defensa, riesgos sociales, gobernanza

Inteligência Artificial na defesa: um exame bioético dos riscos sociais e da governança

Resumo: Este artigo explora os riscos sociais que podem surgir da integração de inteligência artificial (IA) em inovações de defesa e examina como o princípio da responsabilização pode ser alcançado nesse contexto. Utilizando uma abordagem descritiva-analítica qualitativa, baseada em relatórios internacionais e estudos anteriores, o artigo identifica quatro domínios críticos de risco social: ameaças à vida humana, uso indevido de terroristas, guerra cibernética e eletrônica, e falta de responsabilização por ações cometidas por sistemas autônomos ou semi-autônomos. Para enfrentar esses desafios, é proposto um quadro de responsabilidade em três níveis —internacional, institucional e sociocultural— que conecta padrões éticos globais com práticas de governança local e valores morais. O estudo argumenta que inovação responsável em tecnologias de defesa alimentadas por IA requer governança ética que garanta responsabilidade, supervisão humana e legitimidade social, equilibrando assim imperativos de segurança com confiança pública.

Palavras-chave: Inteligência Artificial, defesa, riscos sociais, governança

[1] Education Department, College of Education, University of Hafr Al-Batin, Hafr Al-Batin City, Saudi Arabia, drshammary@uhb.edu.sa, https://orcid.org/0000-0002-0364-7729

Introduction

The world is experiencing an unprecedented acceleration in the development of artificial intelligence (AI) technologies, which have become a defining feature of the contemporary era and an integral component of economic, social, security, and defense domains. However, this rapid advancement is accompanied by serious risks. Omohundro*(1)* warns that intelligent systems—even those designed with relatively simple objectives—may engage in harmful behaviors if not adequately constrained. These risks are further intensified within the defense sector, where Bostrom*(2)* and Altmann and Sauer*(3)* caution that AI may be employed in the development of autonomous weapons or in the conduct of military operations that bypass international norms, thereby threatening both national and global security. In this context, Thurnher*(4)* argues that the deployment of such systems in armed conflicts raises profound legal and ethical dilemmas that undermine their legitimacy.

Beyond these concerns lies a more complex challenge known as **responsibility gaps**. Intelligent systems may generate outcomes that are difficult to predict, control, or fully comprehend, raising fundamental questions regarding accountability. Several scholars describe this phenomenon as an ethical and legal vacuum that threatens core principles of justice and responsibility*(5-8)*. Consequently, the risks posed by AI extend beyond immediate security threats to challenge the moral foundations of social and political systems.

In response to these challenges, international institutions have begun to develop global regulatory frameworks. Most notably, UNESCO adopted the *Recommendation on the Ethics of Artificial Intelligence* in 2021*(9)*, the first international instrument aimed at guiding AI development in accordance with human rights, justice, transparency, and human oversight. Similarly, several scholars emphasize the need to reconcile technological progress with the ethical foundations of society*(10,11)*. These initiatives align with the warnings of physicist Stephen Hawking, who cautioned that the development of advanced AI without ethical, legal, and human safeguards could pose an existential threat to humanity*(12,13)*.

This study does not focus on a specific military AI technology; rather, it addresses risks wherever they arise across diverse AI-enabled defense applications, given their shared ethical and security challenges. It seeks to fill a significant gap in the literature by offering a comprehensive critical analysis of the risks associated with AI-driven defense innovations, particularly in light of intensifying strategic competition among states and growing concerns over the lack of accountability and governance resulting from increased system autonomy*(14,15)*.

The study is grounded in a theoretical framework that links the risks of AI-based defense innovations to the principle of **responsibility**, understood through ethical values as the primary reference for regulation. Existing literature indicates that defense-related AI systems may threaten human life, facilitate terrorism, expand cyber warfare, and generate profound responsibility gaps that complicate the attribution of blame and accountability*(2-6)*. Such risks cannot be adequately addressed through technical or legal measures alone; rather, they require a moral and ethical framework grounded in values such as human dignity, justice, transparency, and fairness. These values may be operationalized at multiple levels: the international level through treaties and ethical frameworks*(9)*, the institutional and regulatory level through oversight and legislation*(16),* and the socio-cultural level through community engagement in shaping the values governing defense applications*(17,18)*.

Accordingly, this study proceeds from the assumption that **ethical responsibility** constitutes the unifying thread between AI-related risks and defense innovations, and that effective regulation of such innovations can only be achieved by aligning technological advancement with shared human values. In light of this premise, the paper seeks to address the following research questions:

What societal risks arise from AI-driven defense innovations? How can the principle of responsibility be achieved in AI-driven defense innovations?

Responsibility Gaps in AI Systems

In this study, responsibility is defined as the relationship between an agent, their actions, and the

resulting consequences*(19,20)*. A growing body of literature has examined responsibility gaps in both military and non-military AI systems. Hellström*(21)* argues that automation weakens direct human accountability, creating a legal and ethical void, while Noorman and Johnson*(22)* maintain that the overlapping roles of humans and machines complicate the attribution of individual or collective responsibility. Purves, Jenkins, and Strawser *(23)* warn that delegating combat decisions to autonomous systems threatens long-standing principles of justice in warfare. Leveringhaus*(24)* further emphasizes the responsibility of political and military leadership to establish strict safeguards prior to the deployment of such systems.

Himmelreich*(25)* extends this debate by asserting that responsibility gaps are not merely legal but also ethical, as the inability to identify a clearly accountable agent undermines the legitimacy of future warfare. In a similar vein, Danaher*(19)* introduces the concept of the **Retribution Gap**, which arises when no moral or legal offender can be clearly identified following harm caused by intelligent systems, thereby eroding trust in justice. Königs*(14)* adds that the multiplicity of actors involved in AI systems—such as developers, manufacturers, vendors, and users—further complicates the accurate distribution of responsibility. Sparrow*(26)* observes that ethical controversy surrounding Autonomous Weapon Systems (AWS) has intensified since the 2012 *Human Rights Watch* report and the subsequent international campaign to ban such systems, while noting that limited moral permissibility may exist in highly constrained contexts.

By contrast, Königs*(14)* adopts a more skeptical stance toward the responsibility gap thesis, arguing that such gaps may be inevitable in theory but insufficiently defined in practice. He contends that concerns about responsibility gaps are often overstated, as there are contexts in which they are unlikely to arise. This position aligns with several studies that question either the inevitability or the significance of responsibility gaps*(27-32)*. These scholars argue that responsibility gaps emerge only in cases where harm cannot be attributed to negligence, malice, or wrongful intent, and that negligent use of autonomous systems remains morally blameworthy.

This debate reveals a fundamental dilemma in defending the concept of responsibility gaps: either one must assume that system autonomy exempts humans from responsibility even in cases of negligence—which is implausible—or accept that responsibility generally remains attributable to human agents, rendering responsibility gaps rare and difficult to identify. It is also notable that much of the literature has focused primarily on the negative dimension of responsibility*(33,34)*. More recent work, however, has expanded the discussion to include the positive dimension of responsibility, such as attributing credit for beneficial AI *outcomes(35-38)*. Nevertheless, the present study limits its scope to negative responsibility due to its direct relevance to the risks under consideration.

Literature Review

The deep integration of AI into various aspects of human life has generated significant academic interest, particularly with regard to its implications for security and societal stability. While a substantial portion of the literature addresses AI-related risks to ethical and religious values in broad terms, comparatively fewer studies focus specifically on security, defense, and warfare-related risks, especially those emerging alongside rapid advances in autonomous and electronic capabilities.

From a defense perspective, Sparrow*(6)* and Thurnher*(4)* argue that autonomous weapons pose a direct threat to human life by undermining the moral and legal foundations of warfare. Bostrom*(2)* and Altmann and Sauer*(3)* warn that AI may be exploited to facilitate terrorism and illicit activities, thereby amplifying threats to global security. Johnson *(39)* and Santoni de Sio and Mecacci*(40)* highlight the expansion of cyber and electronic warfare, emphasizing that unclear accountability increases the likelihood of irresponsible use. Danaher *(19),* Himmelreich*(25)*, and Nyholm*(7)* further stress the complexity of legal and moral responsibility in AI-enabled defense contexts, while Buhmann and Fieseler*(41)* propose frameworks for responsible AI innovation.

Beyond defense, scholars note that accelerated digitalization places increasing pressure on public values, underscoring the need for proactive governance mechanisms*(15,42-44).* Religious and ethi-

cal studies suggest that AI may reshape spiritual practices and cultural authority, potentially weakening traditional religious structures while also enabling new forms of dialogue*(45-49).* Moreover, societal trust in AI systems is closely tied to transparency, human oversight, and adherence to ethical values*(47,9).*

Overall, the literature indicates that most studies have addressed AI either from technical perspectives or from ethical and social viewpoints disconnected from defense applications. Although defense-related risks have been acknowledged, the relationship between such risks and ethical values as a foundation for responsibility-building remains insufficiently articulated. Accordingly, this study seeks to address this gap by examining four key defense-related risks-—threats to human life, terrorism, cyber warfare, and responsibility gaps—and by proposing practical levels for operationalizing responsibility within an integrated ethical and defense-oriented framework.

Methodology

This paper adopts a qualitative descriptive-analytical approach through a systematic examination of relevant literature, previous studies, and international reports on artificial intelligence and its associated risks—particularly in the domains of defense and security innovations. The literature was selected based on specific criteria, including recency of publication (with emphasis on works published after 2010), direct relevance to defense-related AI or ethical and legal responsibility dimensions, and citation frequency, in addition to sources issued by authoritative international institutions such as the European Union, UNESCO, and the International Committee of the Red Cross.

The search process was conducted across major academic databases such as *Scopus*, *Web of Science*, *Springer*, and *IEEE Xplore*, in addition to reports issued by specialized research centers. The paper identifies and synthesizes the most prominent risks extracted from this analysis, with particular focus on the dangers associated with AI-driven defense innovation. Beyond merely identifying risks, the study advances toward proposing mechanisms for operationalizing responsibility in such innovations across three integrated levels: the international level, the institutional—regulatory level, and the socio-cultural level.

Study Limitations

This paper is limited to addressing *negative responsibility* associated with the defensive risks of artificial intelligence and does not examine *positive responsibility* or cases in which intelligent technologies are credited with success. The analysis is confined to available literature without incorporating field-based or technical experimental research, which aligns with the theoretical nature of this study that aims to construct an initial analytical framework—one that may serve as a foundation for future research with deeper empirical or technical engagement. Additionally, the paper does not delve into the algorithmic or engineering intricacies of AI systems; rather, its scope is restricted to identifying defense-related risks and outlining mechanisms for enforcing responsibility within this domain.

Results and Discussion

The Societal Risks of AI-Driven Defense Innovation

Sparrow*(26)* argues that the development of autonomous weapons systems is nearly inevitable, driven by a combination of military and political incentives that encourage the adoption of technologies capable of independently performing combat operations—including the use of lethal force without direct human intervention. Project Maven in the United States is considered one of the most prominent initiatives aimed at accelerating the integration of artificial intelligence and machine learning into U.S. military operations *(50).* Scharre *(51)* employs the term *race to the bottom* to describe the intense competition among states that may lead them to lower safety standards or disregard ethical and regulatory frameworks in pursuit of rapid armament and strategic advantage. Based on the qualitative analytical approach adopted in this study, the most salient risks associated with AI-based defense innovation can be identified as follows:

Threats to Human Life

Sacred traditions have long affirmed the inviolability of human life, considering any violation thereof an act of moral and social corruption. Respect for human life is foundational to social security and psychological stability and is recognized as the most fundamental of human rights. All modern states and international bodies uphold this principle without discrimination on the basis of religion, race, or language. Article 3 of the Universal Declaration of Human Rights explicitly states that "*Everyone has the right to life, liberty and security of person"(52).*

Despite the developmental opportunities offered by artificial intelligence, it has increasingly become a source of existential concern among experts, who warn that it may evolve into a direct threat to humanity. Several prominent technology leaders have voiced such fears. Elon Musk, for instance, cautioned that AI could potentially ignite a future world war*(53),* while Stephen Hawking warned that developing full artificial intelligence without adequate ethical and legal safeguards could lead to human extinction*(54).*

Strategic actions taken by global powers lend further credibility to these warnings. China, for example, has announced its intention to become the world leader in AI by 2030, capturing nearly 60% of global investments between 2013 and 2018*(55).* Russia aims to convert 30% of its military equipment into robotic systems by 2025*(56),* while the United States adopted a comprehensive strategy in 2018 to accelerate AI integration across all branches of its armed forces*(57).*

Al Dehshan*(58)* notes that such rapid military adoption of autonomous systems—such as robots and unmanned aerial vehicles—increases the likelihood of catastrophic errors beyond human control. Similarly, Sparrow *(6)* contends that delegating combat decisions to machines raises profound ethical and legal dilemmas, particularly regarding accountability for war crimes, thereby undermining legal deterrence and fundamentally transforming the nature and conduct of warfare.

Expert estimates further suggest that even a small probability of achieving human-level artificial general intelligence—estimated at only 3%–is sufficient to classify its risks as existentially significant. Some forecasts indicate that this probability may exceed 50% by 2050*(47).* Moreover, several legal studies argue that autonomous weapons lack the full capacity to distinguish between civilians and combatants, or between healthy and wounded soldiers, rendering their deployment—especially in urban settings—highly controversial under international humanitarian law*(4).*

Terrorist Exploitation

This paper contributes to expanding the discussion surrounding the growing potential for artificial intelligence (AI) to be exploited in terrorist operations by linking this dimension to the risks of defense innovation—an analytical perspective that adds depth to existing literature addressing this issue (2,61). As extremist groups increasingly adopt modern technologies for propaganda, recruitment, and operational execution, it is no longer plausible to dismiss the possibility that AI may eventually be weaponized to facilitate more precise and destructive forms of terrorism. This likelihood becomes even more realistic amid escalating geopolitical tensions and the emergence of new conflict zones, which often provide fertile ground for radical ideologies and the formation of extremist cells.

Recent literature highlights that AI's exceptional capabilities in simulation, strategic planning, cyber intrusion, and deepfake manipulation make its weaponization by terrorist groups an imminent threat worthy of serious attention. Extremist actors have already begun experimenting with such tools for propaganda and operational planning. Ahmed *(59)* confirms that contemporary extremism has increasingly benefitted from the digital environment to amplify its activities. Abdel Wahab et al.*(60)* document how ISIS utilized explosive-laden drones to carry out attacks and targeted assassinations, including the attempted strike on Venezuelan President Nicolás Maduro during a military parade in 2018. Bullen*(61)* further warns of the dangers posed by drone systems enhanced with facial recognition technologies for conducting precision assassinations. In a more recent incident, former soldier Matthew Levinson was convicted of using AI technologies—includ-

ing ChatGPT—to plan the detonation of a Tesla truck outside the Trump Hotel in Las Vegas*(62).*

Thus, AI-enabled terrorism is no longer a hypothetical concern but an escalating threat advancing in parallel with the rapid evolution of defense innovations. This underscores the importance of the present paper in drawing attention to a highly sensitive security dimension—one that demands urgent legislative, ethical, and security interventions to prevent the diversion of AI technologies toward terrorist exploitation, especially amid intensifying strategic competition among global powers and increasing fragility in international security.

Cyber and electronic warfare

The literature warns that responsibility gaps in autonomous systems extend beyond conventional military domains into the realm of cyber and electronic warfare. Sparrow *(6)* cautions that delegating decisions to autonomous systems may lead to consequences that fall outside the scope of traditional war crimes. Johnson*(39)* argues that assigning responsibility to individuals helps maintain pressure on developers to ensure system safety and reliability—highlighting a serious risk whereby detachment from responsibility may incentivize developers toward profit-driven neglect. Santoni de Sio and Mecacci*(40)* similarly warn that the ease with which designers and operators can evade blame reduces incentives to prevent harmful outcomes, posing a particularly acute risk in cyberattacks where intelligent systems are capable of obfuscation and large-scale damage without a clearly identifiable perpetrator.

The threat of AI-driven cyber warfare has expanded far beyond the conventional use of autonomous systems. Research now points to the development of *automated hacking systems* capable of independently discovering vulnerabilities, generating malware, and executing integrated attacks without human intervention—complicating legal and ethical accountability and rendering damage more opaque and difficult to trace*(63).* This risk is further exacerbated by the rise of *adversarial attacks* and *model poisoning*, where attackers manipulate data or train systems toward erroneous behavior that leads to catastrophic decisions. Recent reports indicate that such attacks may be used to disable sensitive defense or surveillance infrastructure, including energy, water, and transportation networks—potentially resulting in severe human and material losses*(64).*

Moreover, the informational dimension represents one of the most dangerous outputs of AI in cyber warfare. Deepfake technologies can now be deployed to launch large-scale disinformation campaigns, incite extremist rhetoric, and disrupt political and social processes. What makes this phenomenon particularly dangerous is its ability to undermine trust in institutions and inflame internal divisions—without the need for any direct military confrontation*(65)*. Compounding the issue is the *black box problem* inherent in AI systems: many models operate as opaque structures whose decision-making processes are difficult to interpret even by their creators. This opacity complicates the attribution of fault or responsibility when AI systems are weaponized in cyber warfare—especially when the resulting harm is non-physical, such as media manipulation or societal mistrust*(66)*.

Additionally, commercial incentives and the competitive race among global powers to develop AI-enhanced cyber capabilities further amplify the risks. Under mounting economic and political pressure, safety standards and ethical controls may be relaxed—leading to the production of advanced offensive systems without adequate oversight or transparency. This scenario foreshadows an unprecedented cyber arms race that could drastically increase the likelihood of large-scale disasters that are difficult to contain*(67).*

Absence of Responsibility for Committing Crimes

Earlier sections discussed responsibility gaps in artificial intelligence as a concept that has sparked significant philosophical and legal debate. However, this issue is no longer merely theoretical—it has become a tangible threat imposed by the rapid acceleration of AI technologies. As intelligent systems grow more complex and their outputs become increasingly unpredictable or difficult to control, the fear of *absent accountability* emerges as one of the most pressing challenges to both hu-

manity and society.

Legal systems across human civilizations rest on the foundational principle that an agent must be held accountable for their actions—whether civilly through compensation or criminally through punishment. Yet, the introduction of intelligent machines into this equation has fundamentally disrupted this paradigm. In autonomous vehicle incidents such as the well-known Elaine Herzberg case*(68*), a pressing question arises: *Who is responsible for the error?* The absent human driver? The programmer? The manufacturing company? Or the AI system itself, which cannot be punished or prosecuted?*(69)*.

This dilemma becomes even more severe when transferred from civilian settings to the defense domain, where mistakes or crimes may have catastrophic consequences. If an autonomous combat system were to carry out an indiscriminate attack on civilians, or make a lethal decision beyond human instruction, *who could be held accountable?* A robot cannot be subjected to criminal sanctions, and assigning full responsibility to developers or military authorities may not always be justifiable—particularly given the complexity of production and operational networks and the multiplicity of actors involved. Here, the contours of what is termed the *responsibility gap* become evident: the absence of a clearly attributable agent capable of bearing moral or legal blame, resulting in an unprecedented ethical and judicial vacuum.

This issue transcends legal debate to constitute an existential threat to human security. As reliance on AI in defense contexts increases, individual mistakes may escalate into collective disasters—endangering lives, destabilizing global peace, and undermining fundamental principles of justice and humanity. With the continued advancement of these technologies, the danger posed by crimes committed under conditions of absent accountability will only expand across political, military, economic, and social domains.

Accordingly, there is an urgent need to develop new ethical and legislative frameworks that ensure fair attribution of responsibility—whether through holding political and military leaders accountable, establishing legal mechanisms that assign partial liability to corporations and developers, or adopting international standards that reanchor defense innovations within the bounds of international humanitarian law. Failure to establish such frameworks would render the responsibility gap a *ticking time bomb*, exposing humanity to unprecedented risks in its history.

Achieving the Principle of Responsibility in AI-Driven Defense Innovation

While some voices emphasize the potential of artificial intelligence—particularly in Autonomous Weapon Systems (AWS)—to eventually exhibit greater adherence to ethical norms in warfare than humans themselves, as suggested by Arkin*(70),* other scholars express profound concerns regarding such capabilities. Peters*(71)* argues that the internal factors guiding these systems' decision-making processes remain largely opaque to humans due to their computational complexity, creating uncertainty around their actions. Nevertheless, there appears to be broad consensus on the urgent need for practical and regulatory measures to confront—or at least mitigate—the serious consequences of unsafe AI deployment in defense and security domains. Although completely eliminating these risks may be impossible, continued efforts remain a moral and social obligation shared across all societal institutions, governments, and relevant stakeholders. Each actor bears a role in anticipating risks, proposing adequate safeguards, and developing effective responses should such threats materialize.

Determining who should be responsible for ensuring that AI systems align with human values and objectives—and for safeguarding the principle of responsibility—remains an ongoing debate. Nyholm*(72)* notes that it is not yet clear *who* should be tasked with designing and implementing the necessary frameworks and legislation to guarantee such alignment, reflecting an institutional and ethical gap that has received insufficient attention in current value alignment discourse*(73,74)*. Accordingly, the ambiguity surrounding the actors responsible for establishing such frameworks directly raises the second question of this study: *How can the principle of responsibility be operationalized in AI-driven defense innovations?* To address this, it becomes necessary to analyze the different levels at

which responsibility may be enacted as an entry point for addressing this gap.

The concept of responsibility clearly reflects the reciprocal dynamics among the three core dimensions identified in this study: society, ethics, and AI-based defense innovation. The findings reveal that responsibility serves as the unifying thread connecting these dimensions across multiple levels—namely: the *international level*, which bears responsibility for ensuring that defense innovation adheres to universal standards; the *societal level*, which assumes responsibility for regulating usage and shaping public expectations; and the *institutional-regulatory level*, which is tasked with establishing ethical frameworks, rules, and criteria governing AI deployment. These levels will be discussed as follows:

The international level

The year 2021 witnessed increasing calls from major global powers—including the United States, China, and European Union member states—to establish an international charter defining the ethical principles governing the use of artificial intelligence (AI), grounded in values such as justice, transparency, and respect for human rights *(75).* Parallel to these efforts, significant regional initiatives emerged—most notably the European experience, which established strict legal standards for data protection and privacy, rooted in Western legal traditions that seek to contain technological risks through binding frameworks.

Several countries have since launched national strategies inspired by these principles and adapted to local priorities. A prominent example is the Kingdom of Saudi Arabia's *National Strategy for Data and Artificial Intelligence* (SDAIA, 2023) *(76),* which emphasizes justice, privacy, transparency, and accountability as foundational pillars for responsible AI deployment. In this context, the Political Declaration on Responsible Military Use of Artificial Intelligence, launched by the United States in 2023 with the participation of more than fifty countries, represented an important step toward building international consensus. The declaration focused on establishing guiding principles for the use of autonomous systems and artificial intelligence in defense, most notably: strengthening human oversight, ensuring compliance with international humanitarian law, and avoiding applications that undermine human dignity or increase the risk of uncontrolled military escalation. Although it is not legally binding, the declaration reflects a growing trend toward establishing shared ethical standards that may pave the way for broader international cooperation and, eventually, more binding agreements*(77).*

However, despite the global dominance of such international and Western frameworks, their implementation remains subject to the cultural and religious specificities of each nation. This underscores the need to acknowledge the influence of these contextual factors in shaping ethical standards. Accordingly, distinctive local models have emerged—such as the Islamic perspective—which offers a unique ethical framework for AI that may enrich the global conversation by broadening it to include diverse moral paradigms.

Although international efforts have made significant progress in formulating general ethical guidelines for AI use in civilian and societal domains, the challenge becomes more pronounced when transitioning to the defense sector. AI-driven military innovations raise more urgent questions regarding control and governance due to their direct implications for national security and global peace. This defensive dimension has become a growing concern for many states, particularly in the absence of clear legislation defining the boundaries of responsibility and accountability, highlighting the urgent need for specialized regulatory frameworks.

The European Union's *AI Act* provides an illustrative example of how establishing stringent legal standards—based on risk classification—can help reduce ethical and security threats *(78).* A comparison between key global initiatives reveals notable differences in foundational philosophy: the former adopts a legal-technical Western model prioritizing risk categorization, while the latter seeks value alignment rooted in cultural and religious identity. This divergence indicates that international frameworks, despite their ambition, may fail to fully account for moral plurality—making the inclusion of local contexts essential for ensuring responsibility in defense innovation.

Table 1. Comparative Overview of Prominent International and National AI Governance Initiatives

Initiative	Authority / Year	Primary Focus Area	Strengths	Limitations / Challenges
UNESCO Recommendation on the Ethics of Artificial Intelligence	UNESCO – 2021	Global framework based on justice, transparency, and human rights	First comprehensive international document on ethical dimensions, widely endorsed	Non-binding general principles; limited enforceability, particularly in defense contexts
EU Artificial Intelligence Act (EU AI Act)	European Union – 2024	Classification of AI systems based on risk levels	Legally binding, strict standards, strong focus on technical governance	Primarily centered on European context; insufficient consideration of cultural and religious diversity
National Strategy for Data and Artificial Intelligence (SDAIA)	Saudi Arabia – 2023	Responsible AI for development and security	Incorporates principles of justice, privacy, transparency, and accountability with reference to Islamic values	Limited to domestic scope; requires clearer linkage with international defense governance frameworks
Political Declaration on Responsible Military Use of Artificial Intelligence	United States and Partners – 2023	Regulating military AI and autonomous systems	Introduces political constraints on military AI use; promotes international cooperation	Non-binding; lacks clear enforcement and monitoring mechanisms

To illustrate the strengths and limitations of such initiatives, Table 1 presents a brief comparative overview, highlighting their diversity and the difficulty of formulating a unified global framework to govern military AI applications.

The comparative analysis above illustrates the complexity of achieving responsibility in military AI governance and underscores the need for more comprehensive frameworks that transcend legal and cultural divergences. In response, this paper proposes a set of models that may serve as a robust foundation for guiding international efforts toward more holistic responsibility mechanisms in the defense sector:

Ethical Review and Shared Responsibility Framework for Pre-Deployment Assessment of Defense AI Projects

*Implementation Mechanism***:** Governments should be mandated to establish national regulations requiring all AI-based defense projects to undergo formal ethical review before approval. Moreover, the principle of Shared Responsibility should be enforced between civilian developers of AI systems and the military institutions that deploy them, ensuring that no party is able to evade legal or moral accountability in the event of violations. A practical example of such implementation can be found in the United Kingdom's adoption of *ethical assessments* within its Ministry of Defence governance framework, which evaluates reliability, transparency, justice, and human rights considerations throughout the system›s lifecycle*(79).*

Federal/National Accountability and Oversight Framework with Independent Auditing Mechanisms

*Implementation Mechanism***:** National legislation should require defense institutions and AI system developers to establish independent auditing units or oversight offices responsible for evaluating ethical and legal compliance both before and after system deployment. A practical precedent for this approach is reflected in the *GAO Accountability Framework for Artificial Intelligence(80)*, in which the United States outlined four core pillars of AI accountability concerning governance, data, performance, and monitoring, accompanied by concrete recommendations for impact evaluation and implementation*(80).*

Legally-Binding International Political Declaration on Responsible Military Use of AI

*Implementation Mechanism***:** Existing political declarations should be elevated into binding in-

ternational treaties, rather than remaining voluntary statements, in order to secure both political and legal commitment. International cooperation must be strengthened through cross-border agreements regulating AI use in military operations and curbing arms race dynamics in this domain*(3)*. A practical example is the *Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy*, signed by the United States and several other nations, which seeks to establish an international consensus on responsible deployment and constraints on autonomous weapons systems*(77)*.

Legal Framework Reinforcing the Link Between International Humanitarian Law (IHL) and Military AI Governance

*Implementation Mechanism***:** Standard-setting procedures should stipulate that any deployment of autonomous systems must comply with International Humanitarian Law, supported by auditing mechanisms to verify such compliance. A relevant model has been proposed by Alexander*(81),* who advocates for a governance framework balancing AI alignment with international human rights law and algorithmic accountability principles.

Governmental and Societal Training and Awareness Initiatives to Foster a Shared Understanding of Responsibility

Implementation Mechanism: Mandatory training programs should be introduced for defense personnel, legislative bodies, and innovation teams to enhance ethical awareness and deepen understanding of technical risks. The *Geneva Academy Report on Artificial Intelligence and Related Technologies in Military Decision-Making* highlights the importance of empowering defense decision-makers to understand the capabilities and limitations of AI Decision Support Systems, while educating military users on ethical issues concerning bias, prediction, and explainability*(82)*.

The socio-cultural level

The socio-cultural dimension demonstrates that achieving responsibility in AI-driven defense innovations is not limited to regulating laws and policies alone; rather, it depends on the extent to which societies are able to actively participate in shaping the values that guide such technologies. Cath*(17)* and Coeckelbergh*(18)* affirm that integrating cultural and religious dimensions into AI development enhances justice and prevents marginalization, while Müller and Bostrom*(47)* argue that societal trust is fundamentally linked to clear human oversight and adherence to ethical values. In this regard, Al Dehshan*(58)* stresses that one of the European Union's foremost requirements is the establishment of monitoring and supervisory mechanisms that ensure intelligent systems never exceed the bounds of human control—an essential foundation for building public trust in such technologies. Stahl et al.*(42)* likewise caution that the accelerating wave of digitalization is placing increasing pressure on public values, necessitating a proactive approach grounded in dialogue and informed policymaking.

At the global level, the Asilomar AI Conference (Future of Life Institute, 2017) produced 23 guiding ethical principles, most notably the preservation of human control, ensuring transparency, upholding accountability, and preventing an arms race in autonomous weapons. These principles are intended primarily to reassure societies that technology exists in their service—not at their expense*(83)*. In the same vein, Sparrow*(26)* warns that militarized competition may hasten the development of Autonomous Weapon Systems (AWS), in direct conflict with the ethical understanding of human dignity. Public opinion surveys further reinforce this stance, revealing widespread societal rejection of granting robots the authority to kill*(84)*. UNESCO*(9)* similarly emphasizes that involving communities in AI-related deliberations enhances policy acceptance and reduces cultural misinterpretation.

However, Gabriel*(73)* and Nyholm*(72)* raise an important concern: value alignment is neither straightforward nor easily achieved. Determining *which* values, goals, or interests AI systems should embody—and *whose* values should prevail in the face of disagreement—introduces the risk of power struggles in which certain groups impose their values at the expense of others. This presents profound socio-cultural risks that demand more inclusive approaches to ensure responsible use. To contextualize these perspectives comparatively, Table 2 below presents key initiatives and stud-

ies addressing the socio-cultural dimension of AI in defense, outlining their focus areas, strengths, limitations, and implications for responsibility.

The comparison in Table 2 indicates that achieving responsibility in the socio-cultural dimension requires more than regulatory frameworks; it depends on enabling societies to take part in shaping values and standards, ensuring effective human oversight, and respecting cultural and religious diversity. However, the primary challenge lies in translating these principles from social acceptance into institutional and legislative commitments capable of governing defense innovations in practice. Accordingly, this paper proposes the following guiding principles to support the operationalization of responsibility at the socio-cultural level:

*Structured Community Participation***:** Involving experts, religious leaders, and civil society organizations in policymaking and in reviewing AI defense applications to assess their impact on shared ethical and religious values.

*Value and Cultural Diversity in Design***:** Adopting design methodologies that account for religious and cultural plurality from the early stages of system development to reduce bias and ensure fairness.

*Effective Human Oversight***:** Reinforcing the human role as supervisor and decision-maker across all stages of the AI system lifecycle to strengthen trust and transparency.

*Public Ethical Audits***:** Conducting regular assessments of AI systems' societal and moral impact and publishing results to ensure accountability.

*Ethical and Cultural Education***:** Integrating religious and moral dimensions into educational and training programs to foster societal understanding of emerging technologies and prevent cultural misinterpretation.

The Institutional and Regulatory Level

The institutional and regulatory dimension represents one of the most critical levels in addressing the risks of AI-driven defense innovations. Ethical principles and general declarations are insufficient unless translated into concrete structures and actionable mechanisms capable of mitigating real-world harms. In this context, Project Maven illustrates the ethical and institutional challenges

Table 2. Comparative Overview of Key Approaches in the Socio-Cultural Dimension of AI in Defense

Initiative / Study	Primary Focus	Strengths	Limitations	Implication for Responsibility
Cath (2018)	Linking AI development to social, cultural, and religious norms	Enhances justice and prevents marginalization	Lacks practical implementation mechanisms	Affirms that respecting social norms is essential for societal responsibility
Coeckelbergh (2020)	Emphasizing cultural and religious diversity	Reduces conflict and misunderstanding	Remains largely theoretical	Highlights value pluralism as part of societal responsibility
Müller & Bostrom (2016)	Building public trust through human oversight	Links responsibility to transparency and human control	Does not address direct military challenges	Human oversight as a core pillar of responsible governance
UNESCO (2021)	Community participation in policymaking	Strengthens social legitimacy and acceptance	Cultural variation complicates application	Public participation reinforces ethical accountability
Sparrow (2016) & Charli (2013)	Ethical objection to autonomous killing and public rejection of robot use of lethal force	Reveals global societal stance against human dignity violations	Difficulty translating moral stances into legal obligations	Societal responsibility manifests in rejecting violations of human dignity

of adopting AI in defense. It faced internal resistance from Google employees who rejected the use of the company's technologies for military purposes, ultimately leading to the annulment of the contract. This example highlights the complexity of the relationship between civilian and military actors and the necessity of establishing clear institutional frameworks that align defense innovations with societal and ethical values*(50)*. Similarly, analyses of national AI policies reveal that most countries focus on education, technology, and government investment, while ethical and regulatory dimensions—particularly the design of responsible algorithms and institutional accountability mechanisms—remain marginalized and underdeveloped, underscoring the need for more comprehensive and inclusive governance approaches *(85)*. Abdelrazek*(13)* argues that the existence of unified technical and ethical standards among manufacturers, owners, and end-users contributes to ensuring safety, justice, and transparency. However, the current reality reveals serious shortcomings. Calo et al.*(16)* emphasize that the existing legal and regulatory infrastructure remains inadequate to keep pace with the rapid evolution of AI and the rise of autonomous defense systems.

Although the European experience in data protection—most notably through the General Data Protection Regulation (GDPR)—marked significant progress in addressing risks related to profiling, it remains insufficient in the defense context, as it does not adequately cover fundamental issues such as discrimination, human dignity, and loss of human control. Similarly, the efforts of the European Data Protection Supervisor (EDPS, 2015, 2016; EU, 2016)*(86-88)* have largely remained within the realm of ethical initiatives without evolving into binding institutional policies, while

Table 3. Comparative Overview of Key Institutional–Regulatory Approaches to AI in Defense

Initiative / Study	Primary Focus	Strengths	Limitations	Implication for Responsibility
Abdelrazek (2024)	Need for technical and ethical standards across manufacturers, owners, and users	Highlights shared responsibility among all actors	Lacks binding enforcement mechanisms	Institutional responsibility must be distributed rather than centralized
Calo et al. (2016)	Identifying gaps in existing legal and regulatory infrastructure	Clearly exposes deficiencies in current laws	Does not offer concrete solutions	Emphasizes need to modernize regulatory structures
GDPR (EU, 2016)	Data protection and privacy regulation	Global benchmark for oversight	Insufficient to address issues of dignity and discrimination in defense	Privacy protection alone is not enough
EDPS (2015, 2016)	European ethical initiatives for rights protection	Highlights human rights concerns	Not yet translated into binding policies	Institutional responsibility requires enforcement authority
Cummings (2021)	Embedding institutional mechanisms to reduce risks	Links institutional training to risk reduction	Focuses primarily on military context	Responsibility requires operational implementation tools
Thurnher (2013)	Risks of autonomous weapons dependency	Early legal and security warning	Predominantly diagnostic with limited practical planning	Institutional responsibility necessitates continuous review
Mitchell et al. (2019); Gebru et al. (2018)	Regulatory tools such as Model Cards and Datasheets	Promote transparency and system disclosure	Not yet widely adopted in defense	Practical tools for accountability and transparency

public awareness of the risks associated with intelligent armament remains limited.

Cummings*(89)* confirms that embedding such institutional mechanisms reduces risks and promotes responsible use of AI in military operations. Thurnher*(4)* likewise warns that unchecked reliance on autonomous weapons systems raises complex legal and security challenges that necessitate rigorous institutional review. Additionally, recent studies propose regulatory tools such as *Model Cards* and *Datasheets* to enhance transparency and disclose characteristics of intelligent defense systems*(90,91)*. Table 3 Shows the main institutional and regulatory approaches to Al in defense.

To overcome these gaps, this paper proposes the establishment of new institutional structures as follows:

- National Centers for Responsible Defense AI**,** tasked with developing binding standards before approving any military AI system.
- Independent Review Committees to assess security and ethical risks—modeled on medical research ethics boards—to reinforce transparency and accountability.
- Mandatory Training Programs for military leaders and AI developers to raise awareness of security and social implications and reduce the likelihood of irresponsible decision-making.

Addressing these risks cannot remain within the realm of theoretical recommendations; it requires a comprehensive institutional and regulatory framework that combines *ex-ante review*, *continuous oversight*, and *periodic evaluation*, grounded in transparency and accountability as fundamental conditions. Accordingly, integrating this dimension with the international and socio-cultural levels represents the most effective pathway to ensuring that defense innovation in the AI era remains responsible and conducive to societal safety and stability.

The analysis across the three levels reveals that achieving responsibility in AI-driven defense innovation demands an integrative approach that transcends the boundaries of each level in isolation. At the international level, treaties and initiatives seek to establish binding frameworks to prevent arms races and ensure state compliance with humanitarian principles, yet they remain constrained by political interests and power dynamics. The socio-cultural level demonstrates that true legitimacy arises from societal trust and value alignment, and that neglecting cultural context risks ethical fragmentation and normative conflict. Meanwhile, the institutional--regulatory level highlights that legal and ethical frameworks must be operationalized through practical governance mechanisms that guarantee transparency and sustained accountability.

The comparison among these levels shows that the absence of any one dimension creates a gap that undermines responsible AI deployment in defense contexts. Thus, balanced integration across the three levels constitutes the optimal pathway for aligning defense innovation with ethical and humanitarian values while safeguarding societal security and stability.

Conclusion

This study has demonstrated that AI-driven defense innovations generate a set of fundamental risks that cannot be overlooked. First, these innovations may constitute a direct threat to human life, particularly through the proliferation of autonomous weapons systems capable of making lethal decisions beyond human control. Second, the study revealed the potential for AI to be exploited in the Terrorist Exploitation, whether through its use in unlawful operations or as a tool for recruitment, propaganda, and incitement to violence. The third risk lies in the domain of cyber and electronic warfare, where AI technologies open new horizons for large-scale digital attacks that are difficult to predict or contain. Finally, the problem of absent accountability—or what has been termed *responsibility gaps*—emerges, as these systems may produce harmful outcomes without a clearly identifiable agent who can be held justly and transparently accountable.

To address these risks, the study highlights that realizing the principle of responsibility requires an integrated, multi-level approach. At the interna-

tional level, binding treaties and regulatory frameworks are needed to govern the military uses of AI, while also accounting for cultural and religious differences in shaping ethical standards. At the socio-cultural level, engaging local communities and religious institutions in policymaking, strengthening human oversight, and ensuring alignment with shared human values are essential to enhance legitimacy and social trust. At the institutional-regulatory level, specialized oversight bodies—such as independent review boards and national centers for responsible defense AI—must be established, alongside the development of modern transparency tools such as *Datasheets* and *Model Cards*, to ensure continuous accountability and adherence to ethical standards.

Responsibility, as addressed in this study, is not confined to a single dimension but serves as the unifying thread linking society, ethics, and defense innovation. The paper has specifically focused on *negative responsibility*—that is, the accountability of different actors for harms and violations arising from military AI applications—while deliberately excluding positive or forward-looking responsibilities highlighted in other literature.

The findings confirm that ethical values constitute the foundation for building responsibility in confronting the risks of AI-driven defense innovations. At the international level, values such as justice, dignity, transparency, and the protection of human life serve as the bedrock for treaties and conventions regulating these technologies. At the institutional-regulatory level, embedding values into national laws and defense policies ensures that innovations remain subject to a clear ethical framework, preventing impunity. At the socio-cultural level, respecting religious, cultural, and moral diversity is a prerequisite for public trust and acceptance, ensuring that AI is not perceived as a weapon undermining identity and values but as a tool that can be controlled to serve and protect humanity. Hence, any debate on responsible defense innovation cannot succeed without placing ethical values at the core, guiding policymakers, developers, and societies alike.

The study concludes that addressing these risks is not merely a philosophical debate about the boundaries of responsibility, but a complex practical challenge that depends on the willingness of states and institutions to adopt binding and transparent mechanisms. The greatest obstacle, however, may lie in the military and economic interests of major powers that dominate AI defense innovations, which could obstruct binding international agreements—compounded by a lack of trust between states, disparities in technological capabilities, and resistance from commercial actors unwilling to accept constraints on their interests. Nevertheless, the researcher contends that advancing comprehensive regulatory frameworks and ensuring meaningful global engagement remain the most vital pathways to prevent these innovations from evolving into an existential threat to humanity.

Declaration of Conflicting Interests : The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding: The author received no financial support for the research, authorship, and/or publication of this article.
Informed Consent Statement: Not applicable.

References

1. Omohundro S. Autonomous technology and the greater human good. *J Exp Theor Artif Intell*. 2014; 26: 303-15.
2. Bostrom N. *Superintelligence: paths, dangers, strategies*. Oxford: Oxford University Press; 2017.
3. Altmann J, Sauer F. Autonomous weapon systems and strategic stability. *Survival* 2017; 59(5): 117-42.
4. Thurnher JS. Legal implications of autonomous weapon systems. *Int Humanit Leg Stud*. 2013; 4(2): 187-212.
5. Matthias A. The responsibility gap: ascribing responsibility for the actions of learning automata. *Ethics Inf Technol*. 2004; 6(3): 175-83.
6. Sparrow R. Killer robots. *J Appl Philos*. 2007; 24(1): 62-77.
7. Nyholm S. Attributing agency to automated systems: reflections on human-robot collaborations and responsibility-loci. *Sci Eng Ethics*. 2018; 24(4): 1201-19.
8. De Jong R. The retribution-gap and responsibility-loci related to robots and automated technologies: a reply to Nyholm. *Sci Eng Ethics*. 2020; 26(2): 727-35.
9. UNESCO. *Recommendation on the ethics of artificial intelligence*. Paris: UNESCO; 2021.
10. Koos S. Artificial intelligence: science fiction and legal reality. *Malays J Syariah Law* 2018; 6(3): 23-9.
11. Healey K, Woods RH Jr. *Ethics and religion in the age of social media: digital proverbs for responsible citizens*. New York: Routledge; 2020.
12. Danaher J. *Automation and utopia: human flourishing in a world without work*. Cambridge: Harvard University Press; 2019.
13. Abdelrazek A. Ethical risks of artificial intelligence applications: an analytical study. *J Educ Coll*. 2024; 1(137): 323-76.
14. Königs P. Artificial intelligence and responsibility gaps: what is the problem? *Ethics Inf Technol*. 2022; 24: 36.
15. Royakkers L, Timmer J, Kool L. Societal and ethical issues of digitization. *Ethics Inf Technol*. 2018; 20: 127-42.
16. Calo R, Kerr I, Froomkin M, editors. *Robot law*. Cheltenham: Edward Elgar Publishing; 2016.
17. Cath C. Governing artificial intelligence: ethical, legal, and technical opportunities and challenges. *Philos Trans R Soc A Math Phys Eng Sci*. 2018; 376(2133): 20180080.
18. Coeckelbergh M. *AI Ethics*. Cambridge: MIT Press; 2020.
19. Danaher J. Robots, law and the retribution gap. *Ethics Inf Technol*. 2016; 18: 299-309.
20. Vincent NA. A structured taxonomy of responsibility concepts. In: Vincent N, van de Poel I, van den Hoven J, editors. *Moral responsibility*. Dordrecht: Springer; 2011: 15-35.
21. Hellström T. On the moral responsibility of military robots. *Ethics Inf Technol*. 2013; 15(2): 99-107.
22. Noorman M, Johnson D. Negotiating autonomy and responsibility in military robots. *Ethics Inf Technol*. 2014; 16(1): 51-62.
23. Purves D, Jenkins R, Strawser BJ. Autonomous machines, moral judgment, and acting for the right reasons. *Ethical Theory Moral Pract*. 2015; 18(4): 851-72.
24. Leveringhaus A. Autonomous weapons and the ethics of political responsibility. *Int Polit*. 2018; 55(6): 852-70.
25. Himmelreich J. Responsibility for killer robots. *Ethical Theory Moral Pract*. 2019; 22: 731-47.
26. Sparrow R. Robots and respect: assessing the case against autonomous weapon systems. *Ethics Int Aff*. 2016; 30(1): 93-116.
27. Burri T. Machine decision-making and responsibility gaps. In: Beck S, Jørgensen RB, Schmidt AK, editors. *Human rights and technology: the 2030 agenda for sustainable development*. New York: Routledge; 2018: 175-7.
28. Kohler S, Roughley N, Sauer H. Technologically blurred accountability. In: Ulbert C, Finkenbusch P, Sondermann E, Diebel T, editors. *Moral agency and the politics of responsibility*. New York: Routledge; 2018: 51-68.
29. Lauwaert L. Artificial intelligence and responsibility. *AI Soc*. 2021; 36(3): 1001-9.
30. Robillard M. No such thing as killer robots. *J Appl Philos*. 2017; 35(4): 705-17.
31. Simpson TW, Müller VC. Just war and robots› killings. *Philos Q*. 2016; 66(263): 302-22.
32. Tigard DW. There is no techno-responsibility gap. *Philos Technol*. 2021; 34: 589-607.
33. Glavaničová D, Pascucci M. Vicarious liability: a solution to a problem of AI responsibility? *Ethics Inf Technol*. 2022; 24:28.
34. Kiener M. Can we bridge AI›s responsibility gap at will? *Ethical Theory Moral Pract*. 2022; 25: 575-93.
35. Vacek D. Meeting the AI achievement challenge: collective and vicarious achievements. *Ethics Inf Technol*. 2025; 27: 25.
36. Karlan B. Human achievement and artificial intelligence. *Ethics Inf Technol*. 2023; 25(3): 40.

37. Kuchtová A. The incalculability of the generated text. *Philos Technol*. 2024; 37(1): 25.
38. Scripter L. The achievement gap thesis reconsidered: artificial intelligence, automation, and meaningful work. *AI Soc*. 2024; Online First: 1-14.
39. Johnson DG. Technology with no human responsibility? *J Bus Ethics.* 2015; 127(4): 707-15.
40. Santoni de Sio F, Mecacci G. Four responsibility gaps with artificial intelligence: why they matter and how to address them. *Philos Technol*. 2021; 34(4): 1057-84.
41. Buhmann A, Fieseler C. Towards a deliberative framework for responsible innovation in artificial intelligence. *Technol Soc*. 2021; 64: 101475.
42. Stahl BC, Timmermans J, Flick C. Ethics of emerging information and communication technologies: on the implementation of responsible research and innovation. *Sci Public Policy*. 2017; 44(3): 369-81.
43. Kizza JM. *Ethical and social issues in the information age*. London: Springer; 2013.
44. Rönnblom M, Carlsson V. From politics to ethics: transformations in EU policies on digital technology. *Technol Soc*. 2022; 71: 102145.
45. Floridi L. The ethics of artificial intelligence. In: *Oxford handbook of ethics of AI*. Oxford: Oxford University Press; 2018.
46. Smith R, Laird J. *AI and religion: The future of faith in a technological age*. Cambridge: Cambridge University Press; 2019.
47. Müller VC, Bostrom N. Future progress in artificial intelligence: a survey of expert opinion. In: Müller VC, editor. *Fundamental issues of artificial intelligence*. Cham: Springer; 2016: 555-72.
48. Ahmed S, Sumi AA, Aziz NA. Exploring multi-religious perspective of artificial intelligence. *Theol Sci*. 2024; 23(1): 104-28.
49. Tampubolon M, Nadeak B. Artificial intelligence and understanding of religion: a moral perspective. *Int J Multicult Multireligious Underst*. 2024; 11(8): 903-14.
50. Malmio I. Ethics as an enabler and a constraint—Narratives on technology development and artificial intelligence in military affairs through the case of Project Maven. *Technol Soc*. 2023; 72: 102193.
51. Scharre P. *Four battlegrounds: power in the age of artificial intelligence*. New York: W W Norton & Company; 2023.
52. United Nations. *Universal Declaration of Human Rights*, Article 3. New York: UN; 2025.
53. Musk E. *Mark Zuckerberg and Elon Musk have argued about the future of AI*. World Economic Forum 2017 Jul 26. https://www.weforum.org/stories/2017/07/mark-zuckerberg-and-elon-musk-are-arguing-about-the-future-of-ai/
54. Choudhury S, Kumar N. The future of AI: Stephen Hawking›s warning and Elon Musk›s vision. In: *AI and society: emerging trends*. New Delhi: Academic Press; 2022: 45-67.
55. South China Morning Post. China's AI industry gets most funding but lags US in key talent, says report. *South China Morning Post* 2018 Jul 13.
56. Congressional Research Service. *Artificial intelligence and National Security (CRS Report No. R46458)*. Washington, DC: Library of Congress; 2020.
57. United States Department of Defense. *Summary of the 2018 Department of Defense Artificial Intelligence Strategy: Harnessing AI to advance our security and prosperity*. Washington, DC: DoD; 2018.
58. Al Dehshan JA. Humanity›s need for an ethical charter for artificial intelligence applications. *J Educ Innov*. 2019; 10(10): 1-16.
59. Ahmed A. *Digital ownership in the age of artificial intelligence: Challenges of reality and the future*. Berlin: Democratic Arab Center for Strategic, Political & Economic Studies; 2021.
60. Abdel Wahab S, El-Giani I, Yahya S. Opportunities and threats of artificial intelligence in the next ten years. *Trending Events J*. 2018; 27: 1-16.
61. Bullen J. Killer robots which use facial recognition before slaughtering people 'will be devastating to humankind. *Mirror* 2017.
62. Brundage M, Avin S, Clark J, Toner H, Eckersley P, Garfinkel B, et al. *The malicious use of artificial intelligence: forecasting, prevention, and mitigation*. 2018. https://arxiv.org/pdf/1802.07228
63. Associated Press. Man who exploded Tesla Cybertruck outside Trump hotel in Las Vegas used generative AI, police say. *AP News* 2025 Jan 7. https://www.startribune.com/man-who-exploded-tesla-cybertruck-outside-trump-hotel-in-las-vegas-used-generative-ai-police-say/601203058
64. Fowl L, Goldblum M, Chiang PY, Geiping J, Czaja W, Goldstein T. *Adversarial examples make strong poisons*. 2021. https://arxiv.org/abs/2106.10807
65. West DM. *How deepfakes undermine truth in politics*. Washington, DC: Brookings Institution; 2019.
66. Crawford K, Paglen T. Excavating AI: The politics of images in machine learning training sets. *AI Soc*. 2021; 36(4): 1067-78.

67. Allen GC, Chan T. *Artificial intelligence and national security*. Cambridge: Harvard Kennedy School, Belfer Center for Science and International Affairs; 2017.
68. Levin S, Wong JC. Self-driving Uber kills Arizona woman in first fatal crash involving pedestrian. *The Guardian* 2018 Mar 19. https://www.theguardian.com/technology/2018/mar/19/uber-self-driving-car-kills-woman-arizona-tempe
69. Al Toukhi M. AI techniques and electronic risks. *Al Fikr Al Shurti*. 2021; 30(116): 59-100.
70. Arkin RC. The case for ethical autonomy in unmanned systems. *J Mil Ethics*. 2009; 9(4): 332-41.
71. Peters U. Explainable AI lacks regulative reasons: Why AI and human decision-making are not equally opaque. *AI Ethics* 2023; 3: 963-74.
72. Nyholm S. Responsibility gaps, value alignment, and meaningful human control over artificial intelligence. In: Placani A, Broadhead S, editors. *Risk and responsibility in context*. New York: Routledge; 2023: 191-213.
73. Gabriel I. Artificial intelligence, values, and alignment. *Minds Mach*. 2020; 30(3): 411-37.
74. Russell S. *Human compatible: Artificial intelligence and the problem of control*. New York: Viking; 2019.
75. European Parliament. *Artificial Intelligence Act: Regulation (EU) 2024/1689*. Off J Eur Union. 2024. https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
76. Saudi Data & AI Authority (SDAIA). *AI ethics principles* (Version 1.0). Riyadh: SDAIA; 2023.
77. United States Department of State. *Political declaration on responsible military use of artificial intelligence and autonomy*. Washington, DC: Bureau of Arms Control, Deterrence, and Stability; 2023.
78. European Commission. *AI Act enters into force*. European Commission; 2024 Aug 1.
79. Toutoungi A, Klein D. *Ethics and regulation of AI in defence technology: Navigating the legal and moral landscape*. Taylor Wessing — Interface: Defence Tech. 2025 Jul 31.
80. Government Accountability Office (GAO). *Artificial intelligence: An accountability framework for federal agencies and other entities (GAO-21-519SP)*. Washington, DC: GAO; 2021.
81. Alexander P. Reconciling automated weapon systems with algorithmic accountability: An international proposal for AI governance. *Harv Int Law J Online*. 2023; 64(1).
82. International Committee of the Red Cross (ICRC), Geneva Academy. *Artificial intelligence and related technologies in military decision-making*. Geneva: ICRC & Geneva Academy; 2024.
83. Future of Life Institute. *Asilomar Conference on Beneficial AI: Principles for beneficial AI*. Presented at the Asilomar Conference on Beneficial AI. 2017 Jan.
84. Charli C. Public attitudes toward autonomous weapons: a global survey. *J Int Peace Secur*. 2013; 17(4): 1-6.
85. Saheb T, Saheb T. Topical review of artificial intelligence national policies: A mixed method analysis. *Technol Soc*. 2023; 74: 102316.
86. European Data Protection Supervisor (EDPS). *Towards a new digital ethics: Data, dignity and technology* (Opinion 4/2015). Brussels: EDPS; 2015.
87. European Data Protection Supervisor (EDPS). *Opinion 8/2016 on coherent enforcement of fundamental rights in the age of big data*. Brussels: EDPS; 2016.
88. European Union. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). *Off J Eur Union*. 2016; L119: 1-88.
89. Cummings M. Rethinking the maturity of artificial intelligence in safety-critical settings. *AI Mag*. 2021; 42(1): 6-15.
90. Mitchell M, Wu S, Zaldivar A, Barnes P, Vasserman L, Hutchinson B, et al. Model cards for model reporting. In: *Proceedings of the Conference on Fairness, Accountability, and Transparency* (FAT) 2019: 220-9.
91. Gebru T, Morgenstern J, Vecchione B, Vaughan JW, Wallach H, Daumé H 3rd, et al. Datasheets for datasets. *Commun ACM* 2018; 64(12): 86-92.

Received: November 29, 2025
Accepted: December 4, 2025