LUISS

Degree Program in International Relations

Course of Security Policies

The Security Implications of Artificial Intelligence in The Military and Healthcare Sectors: a Comparative Study.

Prof. Carlo Magrassi

SUPERVISOR

Prof. Stefano Za

CO-SUPERVISOR

Elena Figiani - 656352

CANDIDATE

Academic Year 2024/2025

Table of Contents

Acronyms and Abbreviations

ABMS: Advanced Battle Management System

AGI: Artificial General Intelligence

AI: Artificial Intelligence

APT: Advanced Persistent Threat

AWS: Autonomous Weapons Systems

C2: Command and Control

C4: Command, Control, Communication and Computers

C4ISR: Command, Control, Communication, Computers and ISR

CALO: Cognitive Assistant that Learns and Organizes

CCPA: California Consumer Privacy Act

COMPAS: Correctional Offender Management Profiling for Alternative Sanctions

COP: Common Operating Picture

CSO: Composante Spatiale Optique

DARPA: Defense Advanced Research Projects Agency (U.S.)

DDoS: Distributed Denial-of-Service

DICOM: Digital Imaging and Communications in Medicine

DL: Deep Learning

DoD: Department of Defense (U.S.)

EIOS: Epidemic Intelligence from Open Sources

EU: European Union

FCAS: Future Combat Air System

FDA: Food and Drug Administration (U.S.)

GDPR: General Data Protection Regulation (EU)

GOFA: Good Old-Fashioned AI

HAL: Hybrid Assistive Limb

HCP: Healthcare Professional

HER: Electronic Health Record

HIPAA: Health Insurance Portability and Accountability Act (U.S.)

IHL: International Humanitarian Law

IMF: International Monetary Fund

INS: Inertial Navigation Systems

ISR: Intelligence, Surveillance and Reconnaissance

JADC2: Joint All-Domain Command and Control (U.S.)

LAWS: Lethal Autonomous Weapons Systems

LiDAR: Light Detection and Ranging

ML: Machine Learning

MUSIS: MUltinational Space-based Imaging System

NGA: National Geospatial-Intelligence Agency (U.S.)

NLP: Natural Language Processing

OECD: Organisation for Economic Co-operation and Development

OODA: Observe, Orient, Decide, Act

PACS: Picture Archiving and Communication Systems

PHEIC: Public Health Emergency of International Concern

PIPL: Personal Information Protection Law (China)

PLA: People's Liberation Army

PSYOPS: Psychological Operations

RADAR: Radio Detection and Ranging

SMIC: Semiconductor Manufacturing International Corp

TSMC: Taiwan Semiconductor Manufacturing Company

U.K.: United Kingdom

U.S.: United States

UAS: Unmanned Aerial System

UAV: Unmanned Aerial Vehicle

UCAV: Unmanned Combat Aerial Vehicles

UGV: Unmanned Ground Vehicle

UN: United Nations

USV: Uncrewed Surface Vessel

UUV: Unmanned Underwater Vehicle

WHO: World Health Organization

XAI: Explainable Artificial Intelligence

Introduction

Artificial Intelligence (AI) is gradually penetrating an increasing number of sectors of our daily lives and is expected to transform the way economic, industrial, social, and political activities are performed. Smartphones with integrated AI systems, navigations systems and mobile mapping, translators and interpreters, natural language interaction with computers—such as the well-known ChatGPT—and targeted online marketing are only few examples of how AI is entering in our daily life. The applications of AI extend far beyond consumer technologies, spanning trade, finance, education, public services, healthcare, and the military domain. Investments in the sector saw a huge increase in the last years, with an AI market that is expected to reach $4.8 trillion by 2033 (United Nations Conference on Trade and Development, 2025). The transformative potential is such that there is talk of a "Fourth Industrial Revolution," which includes AI, quantum computing, machine learning (ML) and its subset deep learning (DL), big-data analytics, robotics, additive manufacturing, nanotechnology, biotechnology, and digital fabrication. This pervasiveness has attracted growing interest from both private actors and public institutions. Over the past decades, private technological companies have invested heavily in AI research and development, and states are progressively recognizing AI as a critical strategic asset, with direct implications for national security. The perception that leadership in AI may translate into economic superiority, military effectiveness, and geopolitical influence is gaining ground and, consequently, a global race for AI leadership is taking shape. In this evolving landscape, the United States and China stand at the forefront, while the European Union seeks to position itself as a regulatory and normative power in AI governance. AI is increasingly perceived not merely as a technological breakthrough, but as a structural factor capable of influencing strategic competition and global power balances.

At the same time, AI raises fundamental questions that extend beyond technological performance. With its incredible data processing and automation capacities, it challenges existing governance frameworks, modifies decision-making and operational processes, raises ethical concerns, and introduces new vulnerabilities. In this context AI is not only an object of technological development but also a subject of global security debates, which frame it as both an opportunity and a potential threat.

In light of this, the present thesis aims to investigate how the implementation of AI technologies may affect contemporary security architectures. It does so by adopting a broadened conceptualization of security that moves beyond a strictly military or state-centric perspective to

incorporate insights of human security and health security, that emphasize also the protection of individuals, societal stability, and the capacity of systems to anticipate, absorb, and respond to crises. To this aim, two sectors are selected as case studies: the military and healthcare. Although these domains may initially appear distant from one another, both are highly relevant to a security discourse. The military domain does not need many explanations. It represents a traditional and direct component of security analysis, as military capabilities shape states and organizations' defense and deterrence capacities. AI has attracted substantial attention within this sector due to its potential to enhance military effectiveness and provide strategic advantages. In the contemporary geopolitical environment, characterized by the emergence of non-conventional and hybrid forms of conflict—including cyberattacks and information operations that heavily rely on emerging technologies—AI increasingly appears as a defensive necessity. Healthcare, by contrast, has a less immediate but no less significant relationship with security. In fact, health and security are deeply interdependent. The Covid-19 pandemic clearly demonstrated the profound security implications of health crises by revealing their destabilizing effects on societies, the strain placed on state capacity, and the erosion of public trust. These dynamics call for a rethinking of what constitutes a secure environment, bringing greater attention to concepts of human security. In this context, AI is increasingly deployed to address major challenges facing healthcare systems, including rising demand for services, shortages of healthcare professionals, and growing administrative burdens. By optimizing procedures, supporting diagnostics, automating administrative tasks, among others, AI promises to improve healthcare delivery while allowing healthcare professionals to devote more time to patient care, thereby contributing to broader notions of health security.

Much of the existing literature examines AI either as a military innovation, a governance challenge, or a driver of geopolitical competition. The present thesis instead examines how AI structurally reshapes decision-making architectures across diverse security domains. Indeed, AI's transformative impact may lie more in the integration of algorithmic mediation within institutional processes than in enhanced autonomy or lethality. Implementing AI reshapes operational procedures, human roles, and responsibilities, thereby potentially reshaping security architectures in multiple, interconnected, and still-unknown ways. Through an analysis of the two sectors, the aim is to identify the opportunities and broader implications of AI integration. The thesis does not aim to portray AI as either inherently transformative or intrinsically dangerous. Rather, it adopts an analytical approach, recognizing that technological innovation does not determine outcomes independently and that AI's impact ultimately depends on how it is developed, governed, and used.

AI is a tool, and its security implications are shaped by governance choices, normative structures, and strategic interactions among stakeholders.

With this base in mind, the first chapter establishes a conceptual foundation, which is particularly useful considering the definitional ambiguity and number of myths surrounding AI. It introduces the concept of AI by outlining its historical evolution. It then addresses the main definitional challenges, clarifying how AI is understood and employed within the scope of this thesis. Then, the distinctions between AI, machine learning (ML), and deep learning (DL) are clarified. Finally, given the high expectations and widespread concerns surrounding AI, the chapter examines the main misconceptions and myths associated with this technology.

The second chapter analyzes AI implementation in the military sector. It explores the main areas in which AI is currently being implemented, including intelligence analysis, decision-making support, target recognition, semi-autonomous and autonomous vehicles, cybersecurity, information operations, logistics, and training and simulations. Drawing on official governmental sources and academic literature, the chapter presents concrete examples of AI-related projects developed by states and institutions. The analysis then turns to the challenges and ethical dilemmas associated with military AI, engaging with the scholarly debate between more enthusiastic, skeptical, and pragmatic perspectives—these positions provide a useful framework for understanding the possible implications of military AI—and addressing ethical concerns that are particularly prominent in this domain.

The third chapter examines AI applications in the healthcare sector, including administrative assistance, diagnostics, medical robotics, virtual assistance, and drug discovery. It then explores the technical limitations and ethical challenges associated with AI deployment in healthcare. The chapter concludes by discussing the importance of a broader concept of health security and the ways in which AI-enabled healthcare systems can contribute to wider security outcomes.

Finally, the concluding chapter offers a comparative assessment of AI implementation across the two sectors, identifying the shared opportunities and common security challenges. Throughout the thesis, the global race for AI development emerges as a central background dynamic. Particular attention is paid to the United States, the European Union, and China—the major powers in AI development and regulation. These countries represent markedly different

approaches to AI governance and deployment, as well as strategic positioning within the contemporary AI order.

1. Artificial Intelligence

Although this thesis does not intend to delve into the technical specifications of AI systems, a brief overview of what AI is—and how it has evolved—is essential to the discussion, particularly to understand its later implementation in the two analyzed sectors and to try to clarify the widespread confusion surrounding this rapidly advancing technology.

Defining AI is not a simple task. AI is a broad and evolving technology, and no single, precise, and universally accepted definition currently exists—often leading to confusion. This technology has evolved significantly since the second half of the last century. Originally developed in scientific and academic environments, AI experienced extraordinary growth when it attracted the interest of the commercial sector, with major companies such as IBM and, later, Google and Apple taking a keen interest. As major tech companies became interested in the technology, it started spreading to other areas of society. We are now at a turning point in AI development, with increased investment and major advancements in research in just a few years. AI is now used for a wide range of purposes and is increasingly becoming part of our everyday lives.

1.1 The evolution of AI

The development of what we now call AI began in the 1950s, as part of broader research aimed at creating "intelligent machines". The aim of this early research was to develop systems with cognitive skills and some degree of autonomy, capable of performing tasks typically carried out by humans (Johnson, 2021). The idea of intelligent machines was first introduced by Alan Turing in his famous 1950 paper *Computing Machinery and Intelligence*, where raised the possibility that a machine could potentially learn from experience in a manner similar to a child. Turing presented an experiment consisting of an imitation game, in which a computer pretended to be a human being (Sheikh et al., 2023). This would later become known as the Turing test. Nevertheless, the term "Artificial Intelligence" was officially coined in 1956 during the Dartmouth Summer Research Project on Artificial Intelligence—a six-week workshop organized by John McCarthy and attended by several pioneers of the discipline (Johnson, 2021). The participants were convinced that intelligence could also be created outside the human brain.

Successively, throughout the decades research in AI technologies has not been regular. The research experienced cycles of progress followed by periods of stagnation. Decades of innovation and investments in AI technologies occurred in the 1950s and 1960s and are known as AI Summer.

These were periods of great optimism and broad interest in the field. Initially, researchers focused on developing programs that could play board games. Early successes included Christopher Strachey's checkers-playing program and Dietrich Prinz's chess-playing program in 1951. Arthur Samuel's checkers program, started in 1952, learned to play at a strong amateur level, demonstrating computers' ability to learn beyond explicit instructions (Rai, 2024). However, soon the research moved to tackling logical and conceptual problems. Significant breakthroughs included the Logic Theory Machine, which was built to prove Bertrand Russell's logical theorems. Another significant milestone was the General Problem Solver, a program that could, in principle, solve any problem. By translating problems into goals, actions and operators, the system then could reason what the right answer was (Sheikh et al., 2023). By the mid-1960s, students of these pioneers were creating programs capable of proving geometric theorems and successfully completing intelligence tests, math problems, and calculus exams. Daniel Bobrow's STUDENT program, developed in 1964, was one of the first natural language processing systems that solved algebra problems from everyday language. The MACSYMA project, initiated at the Massachusetts Institute of Technology in 1968, was another sophisticated interactive tool designed to address a variety of mathematical problems (Rai, 2024). Nevertheless, these periods of enthusiasm were followed by phases of disillusionment. Due to misjudgments and exaggerations—often overly optimistic expectations about the capabilities of these technologies—researchers were confronted with the difficulties involved in developing such programs. This frequently led to a loss of faith in AI technologies and a reduction in funding. These periods are referred to as the AI winters and occurred during the 1980s (Sheikh et al., 2023).

However, starting from the late 1990s, research in AI began to accelerate, moving from the laboratory into widespread societal application. This era, often described as AI "leaving the lab and entering society", was characterized by a focused pursuit of specific subfields and real-world applications, such as image recognition and medical diagnostics (Sheikh et al., 2023). This period was marked by increasing number of patent grants, growing private investment, the emergence of new business models, and a rise in AI-related employment. In 1997, IBM's chess-playing computer Deep Blue—capable of evaluating 200 million chess positions per second—defeated the world champion Garry Kasparov, marking a milestone in the history of supercomputers and AI systems that can simulate human thinking[1] (Sheikh et al., 2023). Around the same years, the U.S. Defense

[1] Sheikh et al. (2023) note that, although Deep Blue's victory at the time was seen as a breakthrough in demonstrating machine intelligence, in reality, chess is not the highest expression of human intelligence. On the contrary, it is a

Advanced Research Projects Agency (DARPA) launched the CALO project (Cognitive Assistant that Learns and Organizes), which eventually led to Apple's Siri intelligent software assistant. Automated chat systems—or chatbots—also began to be used by organizations for customer support, combining language processing with expert systems to provide prompt assistance. In 2002, another breakthrough occurred with the first DARPA Grand Challenge competition: a race between completely autonomous road vehicles driving from Los Angeles to Las Vegas (Sheikh et al., 2023; Executive Office of the President, 2016).

The AI systems of the early years generally were distinguished between two approaches. The first is what it's called *symbolic AI*, also referred to as ruled-based AI or Good Old-Fashioned AI (GOFA). Within this paradigm computers are based on the representation of the concepts through symbols and operate by encoding clear logical rules—like "if/then" pairing instructions—to deduce how to behave (Sheikh et al., 2023). This type of AI works well with applications that have clear-cut rules and goals, such as playing chess. However, its capacity is limited in environments characterized by uncertainty or variability, where rigid rule structures are insufficient. In recent years, and especially since the 2010s, a second approach acquired more influence, the so-called *connectionist AI*. This approach draws inspiration from the structure of the human brain and utilizes artificial neural networks to learn patterns from data—essentially represents the foundation of modern deep learning (Sheikh et al., 2023). The emergence of such approach coincides with a new wave of AI progress that began around 2010, when AI saw a new explosion of interest in the field. This *AI renaissance* was driven by three interrelated factors. First, there was a significant increase in the computer capabilities: computers became more powerful and simultaneously more affordable, making them widely accessible. Second, the volume of *big data* grew rapidly from sources including e-commerce, businesses, social media, science, and government. Data has become particularly precious—to the point of being considered the new oil (The Economist, 2017). This surge was closely related to the rise of the Internet: as people used Internet more and more, they directly and indirectly generated vast amounts of digital information, therefore exponentially increasing the amount of data available for AI systems to analyze. Third, there were advancements in machine learning (ML) approaches and algorithms. The increase in available training data led to better-trained ML models. For example, tagging friends on Facebook helped train facial recognition algorithms by providing labeled data (Sheikh et al., 2023). Emerging

mathematical problem with very clear rules and a set of alternatives. As such, a chess program is more about making calculations and selecting among predefined options—tasks already widely performed by machines.

from the technological revolution of the 2010s were language models and pre-trained language models, such as GPT-3, able of using knowledge learned from one or more tasks and apply it to new tasks (Roumate, 2024).

1.2 Definitional Challenges

This historical evolution of AI research provides a base for attempting to define AI and to clarify the conceptual ambiguity that still surrounds the term today. Indeed, defining AI remains complex. It is characterized by a wide range of applications, which frequently leads to it being framed as a *general-purpose technology* (Sheikh et al., 2023). Unlike railways for transportation, telephones for communication, or missiles for warfare—which are single-purpose technologies—AI is more comparable to electricity: an *enabling technology*. It is not designed for a single specific purpose; rather, it can be integrated into a broad range of applications and technological systems in civilian as well as military contexts. AI can be understood as a potential capability multiplier for existing technologies (Horowitz, 2018). This characteristic also explains its designation as a *dual-use* technology, originally designed for civilian purposes but easily adaptable for military and defense application, and further complicates the objective of defining.

AI can be defined in various ways depending on the adopted perspective: technical, functional, philosophical, or legal. Some definitions emphasize the replication of intelligent behavior, while others focus on modeling human cognitive processes. Moreover, many definitions risk being either too narrow—overly confining the technology by excluding a variety of possible and potential applications—or too broad—failing to highlight the unique capacities of AI-powered systems (Johnson, 2021). The presence of such a variety in definitions is not due to carelessness but rather reflects the multidisciplinary and constantly evolving nature of the phenomenon itself. Indeed, AI is a technology that encompasses a wide variety of different applications, functions, and purposes and therefore cannot be confined to just a set of process. Moreover, AI is a rapidly evolving technology, and any attempt to define it must take into account not only its current capabilities but also its potential future developments. This is another reason why overly narrow definitions risk becoming outdated or inadequate within a short span of time. Finally, AI aims to simulate something already difficult to frame itself: human intelligence (Sheikh et al., 2023). Some scholars suggest abandoning the frequently used concept of AI as a replica of human intelligence; instead, they propose considering AI as a different kind of entity: one that makes predictions and

performs tasks associated with human cognition, yet lacks personality, consciousness, or emotions. This is because, although the technology may appear to perform typically human activities, it actually functions quite differently (Sheikh et al., 2023).

Johnson (2021) provides a comprehensive definition of AI, describing it as a universal concept that refers to all the systems aimed at improving the performance of automated processes across a wide range of complex tasks including: (i) *perception*: gathering data from sensors, computer vision, cameras, audio, and image processing, and elaborating it in a coherent manner; (ii) *prediction*: through the study of the data gathered, it can elaborate forecasts on the evolution of phenomena; (iii) *reasoning and decision-making*: analyzing information through tasks of problem-solving, searching, planning, and reasoning; (iv) *learning and knowledge representation*: improving analytical performance through data training; (v) *communication*: conducting conversation in the same way people do, through subfields of AI called natural language processing (NLP) and natural language interpretation; (vi) *autonomous* systems and robotics; and (vii) *human–AI collaboration*: where humans define the systems' purpose, goals, and context and the machine address the set of issues. The definition proposed by Johnson (2021) provides a useful framework of the broad AI's capabilities and applications.

When defining AI, it is also possible to consider the legal definitions provided by relevant institutional bodies. The European Union (EU)'s *Artificial Intelligence Act* (AI Act)—the first comprehensive AI regulation, which entered into force in 2024—in Article 3(1) defines an "AI system" as:

> "a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments" (Regulation [EU] 2024/1689, 2024).

This definition presents a structured and technical understanding of AI. It introduces several core features commonly used to describe AI systems. First, AI is described as a machine-based system, which means is a system that relies on both hardware and software to perform tasks. Second, it operates with varying levels of autonomy. Like the EU, most of the definitions do not specify a precise level of autonomy. This may seem vague and leave room for ambiguity: at what level of human intervention can a technology be considered *autonomous*? Broadly speaking, even

a thermostat that automatically records the temperature and adjusts the radiators accordingly could be considered AI; however, it is not. This is a controversial issue as AI exhibits different levels of autonomy from human intervention depending on the task and objective of the model. A system might autonomously analyze data and propose a decision but still require human intervention to implement that decision. Other systems may require human oversight throughout the whole process. Fully autonomous AI systems remain uncommon, partly because academics remain divided about their legal, political, and ethical implications. As a result, the EU have opted for a broader conceptualization to include all present and future degrees of autonomy. Third, AI may exhibit adaptiveness. This feature is emphasized in other definitions as well, such as that of the United Nations (UN) which refers to AI as a "self-learning, adaptive systems" (United Nations, n.d.). Indeed, certain AI systems are capable of learning from data and improving their performance. Finally, AI receives an input—which could be machine- and human-based—and offers an output such as predictions, content, option of action, decision that can influence physical or virtual environments.

Although the definition may appear complex and dense, it is one of the most comprehensive currently available. The United States (U.S.) does not have a single official definition of AI in its regulatory framework. Nevertheless, the U.S.' 15 U.S. Code § 9401 provides a similar definition:

> "a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. Artificial intelligence systems use machine- and human-based inputs to perceive real and virtual environments; abstract such perceptions into models through analysis in an automated manner; and use model inference to formulate options for information or action" (U.S. Code, 2020).

Additionally, the U.S. Department of Defense (DoD) provides a broader, less technical definition:

> "the ability of machines to perform tasks that normally require human intelligence – for example, recognizing patterns, learning from experience, drawing conclusions, making predictions, or taking action – whether digitally or as the smart software behind autonomous physical systems" (U.S. Department of Defense, 2018).

The definitions provided by the Johnson (2021) will serve as the primary reference point for the purposes of this thesis, as it offers an adequate structure for understanding the applications also observed in the sectors analyzed, while recognizing that still no single definition can fully encapsulate AI's diverse and evolving nature.

1.3 Artificial Intelligence, Machine Learning and Deep Learning

To further clarify the nature of AI, some technical distinctions will now be addressed. It is essential to distinguish between two types of AI and between Artificial Intelligence (AI), Machine Learning (ML), and Deep Learning (DL).

One key distinction is the degree of AI capability, which can be better understood as a continuum that ranges from *narrow AI* to *general AI* (AGI). Narrow AI is designed to address specific tasks without performing anything beyond these tasks. Examples include playing strategic games, driving vehicles, and recognizing images, as well as generating new content—such as text, images, audio, video, or code—by learning patterns from large datasets; the latter are commonly referred to as generative AI systems. Narrow AI is the most common form in use today. On the other hand, AGI (also known as *strong AI* or *superintelligent AI*) refers to a system that can demonstrate intelligent behavior comparable to that of a human across the full range of cognitive tasks. Such an algorithm can not only perform narrow tasks but also functionally think for itself and design solutions to a broader class of problems (Horowitz, 2018). Currently, no forms of AGI exist, though research in the field is ongoing. In the middle of this spectrum, there is *transformative AI*, a form that can go beyond a narrow task but still is not capable of achieving superintelligence (Horowitz, 2018).

In addition, within AI, two related terms emerged in the last years, often leading to further confusion: machine learning (ML) and deep learning (DL). The relationship among these three frequently is confused, however this can be visualized as a hierarchy: AI is the broadest category that encompass all, ML is a subset of AI and focuses on learning from data; and DL is a further subset of ML that relies on deep neural networks to model complex patterns in data.

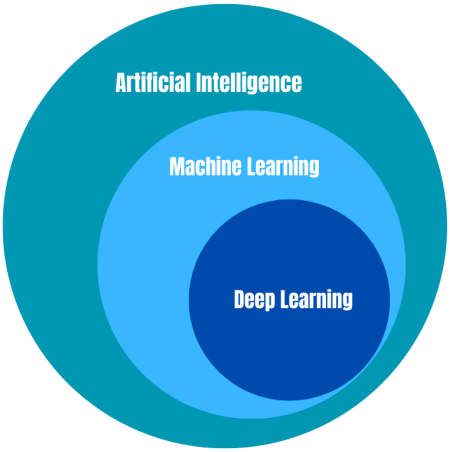

Figure 1- Hierarchical relationship between AI, ML and DL (source: author's elaboration).

Machine Learning (ML) is one of the most prominent and widely adopted technical approach within AI, forming the base of many recent advances and commercial applications of AI. An ML algorithm is a model that begins with a dataset and attempts to derive rules or procedure that explains the data or can predict future data. This form of AI is generally subdivided into three categories: *supervised*, *unsupervised*, and *reinforcement learning* (Kühl et al., 2022; Sheikh et al., 2023). In *supervised* ML, the algorithm is trained on labeled data, meaning it is first trained with input that are paired with the correct output. After learning these associations, the model is tested to see whether is able to predict output for new, unlabeled data. Common application of supervised ML includes spam detection and image classification (Sheikh et al., 2023). In *unsupervised* ML there is no training, and the algorithm scope is to find patterns within the data itself. The model is fed with large amount of unlabeled data in which, without any external instruction, it starts to recognize patterns. This type of ML is usually used when it's not clear what patterns are hidden within the data. In *Reinforcement learning* ML, the algorithm is trained through a process of rewards or penalties for following certain strategies. The algorithm learns to maximize cumulative rewards and develops the optimal strategy (Sheikh et al., 2023). This approach is often used for applications like robotics and game playing.

In recent years, ML technology advanced impressively leading to the birth of the subcategory of deep learning (DL), also known as deep network learning. DL consists in a set of units or "neurons", where each unit combines a set of input values to produce an output. Successively the output is passed on to other neurons downstream. An example of practical application is given by image recognition: a first layer of units analyses the raw data of the image to recognize simple patters in the figure; a second layer of units combines the result of the first layer to recognize patterns-of-patterns; a third layer may combine the results of the second layer, and so on for the analysis of extremely complex, precise patterns in data (Sheikh et al., 2023).

Throughout the thesis, reference will often be made to different systems using the general term "AI," without explicitly specifying the underlying model type—such as AI, ML, or DL—where this distinction is not essential to the analysis.

1.4 Misconceptions and Myths

The recent surge in AI development has given rise to numerous myths surrounding this technology, often distorting public perceptions and generating unrealistic expectations and unfounded fears. Giants of the tech world such as Amazon's CEO Jeff Bezos and Elon Musk have

presented AI as a transformative technology that will radically reshape the world. This dichotomy is reflected in a broader divide between those who view AI as the ultimate technological salvation and those who perceive it as an existential threat to humanity. Although much of this remains far from today's reality, misconceptions surrounding AI persist, particularly in relation to how it functions and the implications of its integration into society.

AI matching human intelligence

One common misconception is that AI will soon equal or surpass human intelligence. This belief is largely driven by ongoing research in AGI, which refers to AI systems that demonstrate intelligent behavior comparable to that of humans. Nevertheless, this narrative is not entirely accurate—the reality is more complex. In fact, research in AGI is still ongoing and there is no consensus on when—or if—it will become a reality. Various companies, such as DeepMind, Vicarious, and Kindred, are working toward developing AGI. However, many scientists remain skeptical, predicting that AGI will not emerge until at least 2060, if at all.

Much of this confusion is fueled by hype from tech companies, which often present advancements in the field as unprecedented breakthroughs. For instance, in the field of autonomous mobility, Otto (a division of Uber) succeeded in creating a self-driving vehicle—specifically a truck for commercial transportation—that autonomously completed a 200 km journey in 2016. This achievement was celebrated as the dawn of fully autonomous vehicles. However, the test was conducted in a carefully managed environment, not on ordinary roads under real-world conditions (Sheikh et al., 2023). The technology used was a form of narrow AI, designed to perform specific tasks under specific conditions. It was not an example of AGI and was not comparable to the cognitive flexibility of a human driver. While Otto's test represented a significant achievement on the path toward autonomous vehicles, suggesting that it marked the arrival of vehicles with human-like driving capabilities was premature. This example illustrates how overpromising commercial applications can distort public understanding of AI's actual capabilities. Tesla, led by Elon Musk, has been claiming for years that it is developing self-driving cars. However, these vehicles will likely not diffuse soon (Elon Musk suggests 2027 as launch date even though optimistic) (Ewing, 2024). In short, the development of AGI and its eventual deployment appear far less straightforward than often portrayed. The notion that AI will soon equal human intelligence is, at present, speculative at best.

AI as malicious

Another fear often found in speculative discourse builds on this misconception and imagines that AI can independently improve itself and become an uncontrollable superintelligence, or worse, turn against humans. These scenarios are often inspired more by science fiction than scientific evidence. Films such as *2001: A Space Odyssey* (1968), in which the supercomputer HAL 9000 tries to kill the protagonists, or *Her* (2013), in which the protagonist falls in love with an AI software, have shaped popular culture imagery, creating the idea of AI being beyond human control. However, these scenarios remain far removed from actual technological capabilities. AI is not an unknown or mystical system. It is designed, built, and trained by humans, who define the algorithm, objectives, and operating logic. For AI to become malevolent and turn against humans, it would need to possess human qualities that it cannot rationally be expected to develop, such as a lust for power, a desire for freedom, jealousy and a fear of death (Sheikh et al., 2023). Research in AGI, as already explained, is still in progress and is far from prospecting this kind of scenario. Fueling fears of a hostile AI uprising is therefore misguided and distracts from more pressing risks. As things stand, the real threat lies elsewhere: the harm caused by AI is less likely to stem from the technology developing evil intentions, and more from human misuse, poor programming, or unintended consequences. For example, a missile system powered by AI could pose a serious risk if poorly programmed or deployed in civilian areas. Therefore, the core issue today is better understood as one of "value alignment": ensuring that AI are developed and trained in accordance with ethical principles and human values, in order to reflect societal values and goals (Sheikh et al., 2023). The real danger does not lie in AI turning evil on its own, but rather in humans using it irresponsibly.

AI as rational

As far as how AI function, common is the belief that AI is rational or possesses something akin to a personality. Interacting with voice assistant like Apple's Siri or Google Assistant, or writing to OpenAI's ChatGPT, can give the impression of engaging with an entity that thinks and reasons like a human. However, this is not the case. Although research in the field is ongoing, current AI models possess limited forms of intelligence, at least far from what we typically associate with human reasoning. At their core, AI systems are basically statistical and mathematical models, trained on massive datasets, that operate based on probabilistic associations and depend

heavily on human oversight, data labeling, and iterative refinement. Behind what seems like intelligent behavior of AI systems there is in fact the labor of people who train, configure, and enable these technologies to function as they do. Large-scale systems such as Amazon's Alexa rely not only on algorithms but also on continuous human involvement to maintain functionality and performance—engineers and system designers who continuously curate the data and optimize responses. Similarly, OpenAI's ChatGPT benefits from a reinforcement learning process that includes human feedback to guide and shape its behavior (Sheikh et al., 2023). Therefore, at least for now, the "intelligence" of such systems remains highly dependent on human input.

The perception of AI as rational is also linked to the concept of AI as a "black box". A black box is a system in which we know the input and the output, but we cannot properly understand the internal processes that translate the input into the output. As a result, AI can appear opaque, undefinable, and lacking transparency—or even giving the illusion of having its own personality. The assumption that AI functions as a black box is largely accurate. This may seem to contradict what has already been said; however, this highlights the inherent complexity of certain AI architectures, particularly ML and DL. While humans design the framework and feed it data, the emergent patterns and internal connections the system forms during its learning process can become so intricate and statistically driven that they are hard for humans to decode (Verbruggen, 2022). Researchers are studying to make these systems more transparent and accessible—the so-called *explainable AI* (XAI). However, this lack of interpretability remains a source of indefinability. This does not imply that AI has a personality or rationality. Nevertheless, it raises concerns about understanding how the system works and, more importantly, whether it is actually safe and trustworthy (Perrigo, 2024).

AI as neutral

Closely tied to the idea of rationality is the assumption that AI is inherently neutral and objective, as it is free from emotions, prejudice, or political convictions. In theory, AI should eliminate the subjective elements that might influence decision-making, thereby ensuring neutral outcomes. Its decisions are expected to result from purely rational, mathematical calculations that evaluate all the parameters of a given situation in an impartial manner. Nevertheless, this assumption has been proven false or not accurate in many real-world cases. For instance, the American Correctional Offender Management Profiling for Alternative Sanctions system (COMPAS) is a software used by US courts to assess the offender's risk of recidivism. Despite

being designed to provide standardized and objective evaluations, it was found that COMPAS overestimated the risk of recidivism in black people and underestimated it in white people (Angwin et al., 2016; Sheikh et al., 2023). AI is not entirely neutral in itself. The neutrality of these systems depends on several factors: the biases and assumptions of developers (the developers' perception and convictions, whether intentional or unintentional, can influence its design), the quality and representativeness of training data (if the data is not entirely free of bias, it influences the algorithm's training), and the societal definitions of equitability encoded into algorithms (which can pose several difficulties). As AI is built and trained by humans, it inevitably reflects human choices—conscious or unconscious. Therefore, assuming AI is inherently neutral is inaccurate.

Having outlined a conceptual framework of what AI is, how it has developed, and having examined the main common assumptions surrounding it, the analysis now will enter into the specifics of the two case studies—the military and healthcare sectors—in order to examine how AI has been implemented in these domains, the principal debates surrounding its adoption, and the key challenges associated with its use.

2. Military Applications of AI

The military domain is currently the sector attracting the greatest interest for the application of AI. States are actively pursuing the integration of AI into military capabilities, recognizing the potential of these technologies to address existing shortcomings and provide innovative capabilities. This recognition is leading to a growing perception that dominance in military AI could translate into a strategic advantage and influence global power balances. This trend unfolds within an international security environment increasingly characterized by non-conventional forms of conflict. Often referred to as hybrid threats, these include cyberattacks targeting critical infrastructures, government, businesses, and citizens, as well as political interface, economic pressure, and psychological operations (PSYOPS), among others. Most of these emerging battlefields heavily rely on advanced technologies—including AI—both for offensive and defensive purposes. Consequently, the effective integration of such technologies has become a strategic necessity for states seeking to respond to these evolving threats and sustain effective security institutions.

In the military arena, there are already many implementations and attempts at optimization through AI systems: from guidance of physical assets like robotic systems, intelligence analysis, surveillance activities, the monitoring of physical sites and online network flow, as well as predictive maintenance of equipment—whether on ships, planes, or in control centers. Such systems expand analytical capabilities and increase automation, enabling machines to support and augment human operations and, in some cases, to perform complex tasks with minimal human assistance or supervision. This chapter will examine the principal AI application in the military domain to better understand the ways and purposes for which this technology is being developed and adopted in this field. It will then address the debate surrounding the potential of AI in warfare––whether transformative or incremental—and finally analyze the key ethical and governance issues associated with the use of AI in military contexts.

2.1 AI Capabilities in the Military Domain

AI's character as general-purpose technology allows AI to enter also every facet of the military domain, with a single AI function being applicable across multiple contexts (Fischer, 2022). With regard to the main AI functionalities deployed in the military domain, several core

categories can be identified. First, AI's ability to process large volumes of data in a short time proves particularly valuable in military fields characterized by information overload—such as military intelligence and information-gathering processes. Second, pattern analysis and forecasting, whereby algorithms detect trends and anomalies in data and generate predictions based on these patterns—for example by identifying changes in the trajectory of a target or tracking the dissemination of online disinformation. Third, planning and optimization, which involves designing sequences of actions or strategies to achieve specific objectives, enable the formulation of alternative courses of action and potential outcomes, thereby supporting strategic analysis and decision-making in mission planning and resource allocation. Fourth, natural language processing (NLP), which enables the identification, analysis, and generation of information from written or spoken communication. Fifth, computer vision, which allows systems to interpret and map elements of the operational environment using sensors such as optical, infrared, or thermal cameras. Sixth, modeling and simulation, which support the testing, optimization, and evaluation of scenarios and strategies in simulated environments. Finally, robotic automation, which refers to the use of software-based automation to perform repetitive and routine tasks (National Security Commission on Artificial Intelligence (NSCAI), 2021). These AI-enabled functionalities are integrated throughout the military in diverse ways and across multiple levels.

2.1.1 Intelligence Analysis

Intelligence analysis refers to the processes of information and data mining, collection, and processing through which information is subsequently conveyed to decision-makers and forms the foundation of their analytical assessments. This activity is fundamental to gaining and maintaining a reliable information level, while also enabling informed and deliberate decision-making. In this context, AI has proven particularly useful, as it enables analysts to manage vast volumes of data from multiple sources and facilitates their analysis. Consequently, intelligence and analytical centers are increasingly implementing AI systems at different levels to enhance efficiency and analytical capacity.

In data mining operations, AI and ML are employed to improve the quality and timeliness of information collection. These technologies enable the analysis of large volumes of both structured and unstructured data within relatively short timeframes. The result is, on the one hand,

multi-source data integration that provides more comprehensive and accurate information. Such activities include, among others, scanning drone footage, analyzing camera feeds, and monitoring online activity. These systems are then used for processing and exploitation, which involves extracting information from the collected data, such as objects classification, recognize behavioral patterns, and potential threats identification. AI can pre-process large volumes of raw data by prioritizing relevant information for transmission and storage while discarding irrelevant data, thereby enhancing overall information usability. For example, AI systems can be used to identify threats such as a military aircraft undertaking bombing runs and to alert civilians in the affected areas (Taddeo, 2024). AI automates processes that would otherwise require significant time and effort from human operators, enabling them to be completed more rapidly and allowing personnel to focus on higher-level analytical and decision-making tasks.

The first and most prominent project in this sense is the Maven Smart System, launched by the U.S. DoD in 2017 and commonly known as Project Maven. Its objective is to relieve human operators tasked with analyzing video footage obtained from unmanned aerial system (UAS) by automatically elaborating the content. The algorithm is able to analyze and fuse huge quantities of data from various sources, identifying anomalies, targets and display information, streamlining the process that was before carried out by human operators (Kahn, 2024). The project initially relied on the open-source Google tool-box TensorFlow. However, following strong opposition from Google employees, the company decided to not extend its contract. The contract was subsequently awarded to Palantir Technologies. In 2023, the DoD transferred control of the project to the National Geospatial-Intelligence Agency (NGA). This technology has been deployed in various real-world operations, including identifying targets in Iraq, Syria and Yemen. During the Russia-Ukraine war, Maven was used to process satellite imagery and relay intelligence to Ukrainian forces (Ibrahim, 2024). In March 2025, NATO finalized a contract with Palantir for the acquisition of the Maven Smart System technology for its intelligence operations (NATO, 2025).

Another noteworthy example is the U.S. system Raven Sentry. The model was developed to predict seemingly sudden episodes of political violence in Afghanistan in 2020. Trained on historical data on violent incidents and using publicly available information—such as weather patterns, social media content, news reports, and commercial satellite imagery—the system was able to predict an attack with a probability of 80–90% in approximately 70% of cases, performing on par with human analysts while operating at a significantly greater speed (The Economist, 2024).

Similar technologies can be observed within the Israeli forces—a leading effort in AI systems development—which have implemented AI technologies in intelligence and surveillance through multiple projects, in particular for subjects' identification purposes. The AI-enabled facial recognition technology developed by AnyVision is used by Israeli forces for identifications in primary two scenarios. The first in Israeli daily surveillance, for the checkpoints where Palestinians pass on their way to work inside Israel, identifying instantly visa holders. The second for intelligence operations: through a network of cameras spread throughout the area, the algorithms perform potential threats track and identification (Dolinko & Antebi, 2024).

Another example is presented by the People's Liberation Army (PLA), which uses AI for real-time surveillance, reconnaissance, and early warning through a range of AI-enabled models, such as the BeiDou satellite navigation system. BeiDou, in addition to its navigational functions, analyzes data and supports threat monitoring and early-warning operations. The scope of Chinese leadership is to "separate the wheat from the chaff" in its intelligence, surveillance and reconnaissance (ISR) inputs, eliminating the useless data and, in particular, overcoming the problems of obfuscation and data corruption that can undermine the validity of an operational picture (Fedasiuk & Weinstein, 2022). Still within the domain of satellite systems, France's program on AI-enhanced systems for the automatic detection of activities at sites of strategic interest also proves highly relevant. The TAIIA project, launched by the French Directorate of Military Intelligence in 2020, provides AI systems to support experts in the search for information and clues within imagery, as well as to improve the analysis of activities at pre-selected geographical sites. The system is fed by operational images from the new CSO (*Composante Spatiale Optique*) observation satellite constellation—French military reconnaissance satellites part of the MUltinational Space-based Imaging System (MUSIS) program[2]. Through its elaboration, TAIIA's system has demonstrated far-reaching operational improvements. By lightening the work of photo interpreters and intelligence operators, it freed them from more mechanical basic operations and enabled to concentrate on higher value tasks, such as improved analysis and producing usable intelligence (Martin & Liversain, 2024).

As illustrated by these projects, AI systems deployed in intelligence analysis, on the one hand, streamline data analysis and enable the evaluation of significantly larger volumes of

[2] The MUSIS is a European program dedicated to sharing imagery form various military satellites through a common generic ground-based center. Belgium, Italy, Germany, Greece, and Spain, along with France, are the other nations part of the program.

information within shorter timeframes. On the other hand, it entails increasing reliance on algorithmic systems for the sensitive task of information selection and assessment. Intelligence evaluation progressively depends on the algorithmic pre-selection of relevant data, meaning that the information reaching human decision-makers is already filtered through machine-generated processes. This represents a delicate shift; ensuring the quality, reliability, and integrity of algorithmic outputs—and that the system does not present bias, error, or manipulation—becomes essential in order to prevent security failures.

2.1.2 Decision-making Support

Another sector living a huge proliferation of AI programs dedicated is Command-and-Control (C2). C2 refers to the work of authority and direction carried out by a structure for the accomplishment of a mission. In recent years, the role of C2 structures has evolved significantly, with the emergence of battle networks and concepts such as "mosaic warfare", which views war as a data-centric, adaptive system, combining platforms, weapons, and sensors to create a more effective and resilient approach[3] (Jensen & Paschkewitz, 2019). Such transformation has progressed to the point that discussions now extend beyond C2 to C4 (command, control, communication and computers), which includes computing to enhance data management and analysis, and C4ISR (Command, Control, Communication, Computers and ISR).[4] In such context, the integration of AI into C2 systems is a key priority for many nations.

In decision-making centers, AI systems are used to elaborate, analyze and centralize multiple data from various sources in order to enhance situational awareness and achieve greater operational interoperability, ultimately enabling more rapid responses (Kahn, 2024). Such systems are employed to fuse information provided by platforms operating on the battlefield and to generate the so-called Common Operating Picture (COP)—a comprehensive image of the operation situation complete of the data coming from all the disposable resources. This process helps resolve redundancies and outstanding discrepancies, providing decision-makers with a clear and coherent

[3] Ad Jensen and Paschkewitz (2019) outline, the concept of "mosaic warfare" first emerged in the U.S. Defense Advanced Research Project Agency (DARPA). The term refers to a strategy of conducting multi-domain maneuvers against adversaries by combining small unmanned systems with existing capabilities in new ways. This approach is intended to provide an inexpensive, rapid, lethal, flexible, and scalable tactic.

[4] An additional distinction is C3 (command, control and communication), which includes communication system to optimize information sharing.

picture of the situation, already coordinated with all available data. In a paper of the European Defence Agency (2020) on the implementation of the EU's defense toolbox, Christian Hedelin—Chief Strategy Officer Saab—explained how AI allows analysts to extract vastly more insight from the data that would otherwise be discarded: "we come from a world where we throw away most of the data just to find certain signal characteristics, to a future where we will be able to squeeze so much more information out of the data that our sensors gather."

The main example of such strategy is the DoD's Joint All-Domain Command and Control (JADC2) strategy, which seeks, also through the use of AI, to promote coherence of efforts and greater interconnectivity across all the military departments—Air Force, Army, Marine Corps, Navy, and Space Force—resulting in advanced command centers. AI represent an important component of such approach as it allows the Joint Force to "sense," "make sense," and "act" on information across the battlespace quickly via a robust network environment (U.S. Department of Defense, 2022). The U.S. departments are implementing the strategy through sector projects, the main ones being the Air Force's Advanced Battle Management System (ABMS), the Navy's Project Overmatch, and the Army's Project Convergence. All three of these make extensive use of AI and ML. The ABMS and the Marine's Overmatch Project focus on the creation of an AI-enabled data network that seamlessly connects the various forces in order to share data, manage information flow and create a unique operating picture. The ABMS is a unit within the Department of Air Force Program dedicated to command, control and battle management, and the project focuses on the connection of air forces, space forces sensors, systems, and weapons (Congressional Research Service, 2022*b*; Air Force Material Command, 2025). Similarly, the Marine's Overmatch Project aims to improve networking and data sharing among ships, aircrafts and weapons using AI technologies. Instead, the U.S. Project Convergence is a continuous experimental program of the US Army aimed at developing the capacity of conducting multidomain operations using new technologies, included AI and ML. The project is composed of various events and demonstrations, culminating in a capstone event. The Project Convergence-21 (PC21), which took place in Arizona in 2021, specifically focused on networking various platforms to accelerate target identification–response decision processes (Lacdan, 2021).

Another relevant example of AI application for data analysis and merge is the platform SitaWare, provided by the Danish company Systematic, which supports situational awareness and military leaders in decision-making on the battlefield by providing instant share of information among operation actors as well as allies. The system is used by several countries, such as most

NATO counties, including Australia, Finland, Germany, Italy, Sweden, the United Kingdom, and the U.S., along with the Danish Army, which, however, uses only a limited-AI-version of it (Systematic, 2023; Borchert et al., 2024). Even the Ukrainian system Delta, developed by the Ukrainian Ministry of Defense, is employed in the current conflict for situational awareness. The system integrates information form a variety of data sources to provide a comprehensive situational picture to all branches and command levels of the Ukrainian military—from ground troops to higher military leaderships—which access it through mobile devices with different level of access permissions (Bondar, 2024). AI implementations in situational awareness, planning optimization and decision-making have been developed and tested also in the context of the AI for Defence (AI4DEF) partnership. AI4DEF was an international project backed and funded by the European Commission as a part of the European Defence Industrial Development Programme. The project, which developed between 2021 and 2024, saw the contribution of 22 partners among companies and research institutions from ten countries[5], including TERMA (acting as consortium leader), Leonardo, and Airbus (European Commission, 2021; Graae, 2022)

With the aim of improving decision-making processes, some systems also provide analyses of courses of action and potential solutions. Indeed, AI can be leveraged to identify patterns and trend, highlighting possible improvements in strategies. Personal assistant technology and gaming systems can suggest tactics, pathways for mission implementation, and options that might have escaped the human eye. AI-algorithms enable simulations and extrapolations of future outcomes across numerous factors, such as battle impact, casualty estimation, and assessments of collateral damage (Grand-Clément, 2023; Cecchini, 2023; Sayler, 2020). The Russian C2 system, ISBU, is planned to analyze the data and propose alternative courses of action based on its assessment of the ground situation (Zysk, 2024). The Chinese War Skull wargaming system, launched in 2020, uses AI to generate complex simulated environments in which to test tactics and discover vulnerabilities, in this way also reducing dependence on large-scale human exercises. At the moment its applicability to real word remains doubtful due to challenges such as unpredictability in complex scenarios (Graham & Singer, 2025). Relevant is also the German project GhostPlay, which develops decision algorithms (Play) for both defensive and offensive tactics though simulation environment (Ghost). The project, funded by the German Bundeswehr, develops AI-enhanced

[5] The ten participating countries are: Cyprus, Denmark, Estonia, France, Germany, Italy, Latvia, Lithuania, Slovenia, and Spain.

defense decision-making algorithms for fast-paced air operations, particularly against swarms of unmanned aerial vehicles (UAVs). The focus of the project is pointing toward a decentralized approach to air defense coordination where individual systems learn to cooperate through emergent behavior rather than relying on central C2 structures (Borchert et al., 2022).

Thus, within C2 structures, AI systems improve interoperability and enable the integration of data into a comprehensive operational framework, allowing for significantly more advanced and rapid analysis. At the same time, this translates into a compression of decision cycles and centralization of data flows within interconnected platforms, thereby modifying operational tempo and narrowing deliberative margins. This transformation creates new categories of risk. On the one hand, when the production of analysis and the construction of COPs are entrusted to algorithmic systems, risks of misinterpretation, data distortion, or manipulation of the AI model arise. This may result in misleading assessments and, consequently, inappropriate response actions. Therefore, ensuring the reliability, transparency, and robustness of AI-generated outputs becomes crucial in order to mitigate these risks. On the other hand, while enhanced responsiveness may improve operational effectiveness, excessive speed can intensify escalation dynamics. Accelerated decision-making processes may reduce de-escalation opportunities and facilitate the triggering of rapid "action–reaction" dynamics. In this sense, the integration of AI into C2 structures does not solely optimize information management but has the potential to reshape procedures' speed and control.

2.1.3 Target Recognition and Autonomous Weapons

AI is widely employed to optimize target recognition, one of its primary fields of application. The integration of AI allows human operators to analyze operational data more comprehensively, drawing on information from multiple deployed and connected sensors. These systems enable clearer distinction between allies and adversaries, civilians and combatants. They facilitate the prioritization of threats to optimize resource allocation and minimize human exposure in combat. Such application of AI in target identification ultimately streamlines the entire target recognition and response process.

In the context of the mentioned Project Convergence-21 (PC21), the U.S. Army developed a network of four AI systems aimed at fastening and improving the target identification–response decision process. During this exercise the four AI algorithms—Rainmaker, Prometheus,

FIRESTORM, and SHOT—by interacting were able to accelerate the OODA cycle (Observe, Orient, Decide, Act), downsizing the time between detecting a target and destroying it from 20 minutes to 20 seconds (Alderman, 2021). Even Israel stands out with several programs for threat alert, targeting and decision-making optimization. For instance, the system Fire Factory analyzes datasets, including historical data about previously authorized strike targets, to calculate amounts of ammunition require, propose optimal timelines and target prioritization and allocation (Mimran & Dahan, 2024). It merges the phases of target identification and evaluation of the military capabilities to find the most appropriate response for engage with the target.

When AI systems are applied to a weapon, it is referred to as autonomous weapon systems (AWS), or, if lethal, lethal autonomous weapons systems (LAWS), namely systems that, once started, can identify and engage targets without further human intervention. Thanks to the sensors, software and physical means of which it is provided—such as optical cameras, infrared cameras, hyperspectral and full spectrum imaging, light detection and ranging (LiDAR), inertial navigation systems (INS) and satellite navigation, and radio detection and ranging (RADAR)—the machine is able to sense the environment, process the information and act accordingly (Dahlmann, 2022).

AWS and LAWS seem to be the future of warfare, reshaping drastically how conflicts are fought. Such machines leverage AI for tasks ranging from surveillance to lethal engagement and can operate with varying levels of human oversight. The increased use of AWS and LAWS is likely to coincide with an increase in the use of armed drones and robots in combat, thus shifting to autonomous agents the majority of functions—as is currently happening in the Ukraine–Russia war with drones. Nevertheless, for the time being, such technologies are still in the early stage of developments, because of practical challenges—a high level of technological sophistication is required—but also for ethical and legal reasons. While, on the one hand, these systems may contribute to a reduction in human casualties on the battlefield, issues of decision-making accountability, the risk of malfunctions, and their ethical implications play a decisive role in the broader debate (Dahlmann, 2022; NSCAI, 2021; Sayler, 2020; Grand-Clément, 2023). Technical failures, algorithmic errors, or misinterpretations of data may lead to unintended engagements, excessive or disproportionate use of force, and serious harm to civilian populations. Therefore, rigorous safeguards are required in the development and deployment of AWS and LAWS to ensure their safety, reliability, and full compliance with International Humanitarian Law (IHL).

2.1.4 Semiautonomous and Autonomous Vehicles

AI is implemented in vehicles, including aircraft, drones, ground vehicles, and naval vessels, to enhance their autonomy. The result is increased efficiency in military operations.

AI systems are key in unmanned vehicles as they enable self-governance in complex environments, especially with regard to movement and navigation. Thanks to several sensors, the algorithm collects and merges the data gathered and, in this way, perceives the environment, recognizes obstacles, plans navigation, and even communicates with other vehicles (Sayler, 2020). Their deployment appears particularly useful in reducing human exposure and increasing operational precision in complex and hazardous environments that require a high degree of accuracy (Sayler, 2020). The most prominent ones are aerial vehicles, commonly known as drones but also referred to as unmanned aerial vehicles (UAV), or unmanned combat aerial vehicles (UCAV) if armed. AI is increasingly applied in the development of drones and swarming drone systems. Ongoing experiments aim to develop multiple autonomous vehicles capable of operating in coordinated formations and effectively teaming with human operators. In this regard, Ukraine has emerged as a leading case in the use of drones, having developed extensive operational experience through battlefield deployment during the conflict with Russia. At the same time, several other military establishments are moving to develop similar capabilities. In Europe, for instance, the Eurodrone programme aims to develop UAVs for intelligence, surveillance, and precision strike missions. The project is developed as a joint industrial effort involving Leonardo, Airbus Defence and Space, and Dassault Aviation, and represents a partnership between industry and government supported by France, Germany, Spain, and Italy (Clapp, 2025*a*).

In the field of flight control and piloting AI is leveraged to enhance or replace human pilots in key flight functions. The U.S. Air Force is currently engaged in the testing of AI semi–piloted and piloted aircrafts (Hodges, 2025). The Future Combat Air System (FCAS)—a German-French-Spanish project for the development of a sixth-generation multi-role fighter system, including remotely piloted aircraft and new weapon and communication systems—addresses AI for supporting functions, such as enabling the complex guidance and flight control behavior to navigate unmanned platforms (Borchert et al., 2024). However, unmanned vehicles are also deployed on the ground or under water. For example, the U.S. Navy's Task Force 59.1 is implementing fully autonomous vehicles with AI-enabled surveillance capabilities for maritime surveillance and detection (NAVCENT Public Affairs, 2024).

The diffusion of autonomous and semi-autonomous vehicles signals a shift toward increasingly complex—but at the same time safer—operations. By enabling access to contested or high-risk environments without directly exposing human operators to harm, these systems expand the range of possible missions while reducing immediate personnel vulnerability.

2.1.5 Cybersecurity

AI is widely used to enhance cyber and electronic warfare capabilities. Cyberattacks are increasing considerably, targeting sensitive centers, public institutions, businesses and citizens. This includes a wide range of digital threats, such as data poisoning, phishing attacks[6], and advanced persistent threats (APTs). Attempts of intrusion through the use of malicious network aimed at compromising information availability, integrity, or confidentiality have become increasingly complex and frequent.

In this context, AI innovations have substantially influenced the cyber risk management landscape, giving rise to both offensive and defensive AI strategies. In defensive cybersecurity, AI is employed to detect cyber threats and predict attacks by analyzing network activity and identifying patterns of normal and intrusive behavior. For example, the Security Operation Centre (SOC) of Leonardo in Italy leverages AI-driven systems to protect critical infrastructures from cyber-attacks and the interruption of digital systems that underpin essential public services (Leonardo, 2022). AI systems are trained using techniques such as adversarial training, anomaly detection, and real-time threat analysis. Within cybersecurity operations centers, these systems can automate routine real-time scanning for suspicious activity, significantly increasing the speed and effectiveness of responses to security incidents. Defensive AI systems further improve their robustness against adversarial manipulation by learning from past attacks and continuously adapting their models to counter emerging AI-driven threats more effectively. AI is also used in software testing, enabling the identification of weaknesses and vulnerabilities before systems become operational (Cecchini, 2023; Rashid et al., 2023). Conversely, in the context of offensive cyber operations, AI is leveraged to identify vulnerabilities in adversary networks, evade detection, deploy malware, and amplify the impact of cyberattacks. For instance, autonomous hacking frameworks harness AI algorithms to identify and exploit software vulnerabilities at scale, while

[6] Phishing attacks refers to a type of cyberattack that uses fake emails, text messages, phone calls or websites to trick people into sharing sensitive data and downloading malware (Kosinski, n.d.).

adversarial ML techniques are used to generate corrupted inputs capable of deceiving AI-based network security systems (Swetha et al., 2025).

AI in cybersecurity illustrates one of the most significant shifts in the contemporary security landscape. On the one hand, the rise of hybrid forms of destabilization—particularly AI-enabled cyberattacks—exemplifies the emergence of conflict strategies based on indirect methods aimed at destabilizing states from within by targeting critical infrastructures. On the other hand, the increasing speed of both cyberattacks and defensive responses, enhanced by AI systems, often exceeds the capacity of human-controlled processes. As a consequence, operational procedures evolve toward greater reliance on AI systems, while human operators shift from direct execution to supervisory oversight. This shift, once again, underscores the criticality of ensuring the efficiency, reliability, and quality of such systems.

Even in electronic warfare AI implementation presents crucial benefits. Military operations conducted in all environments—air, land, maritime, space, and cyber—greatly rely on the electromagnetic spectrum. Electronic warfare involves the use of the electromagnetic spectrum to attack the enemy or impede enemy operations, including attacks on radar systems, jamming of communication and navigation systems, electronic masking, and reconnaissance and intelligence gathering, among others (NATO, 2023*a*). In this context, AI implementation significantly enhance the capabilities related to sensing, signal identification, countermeasure execution, and integrated cyber defense. AI analyses the data collected across the full electromagnetic spectrum (which includes visible, ultraviolet, infrared, microwaves, and radio waves) in specific domains to identify elements of interest (Grand-Clément, 2023). AI algorithms are applied in UAS for tasks such as radar signal identification. Moreover, AI Is exploited for tiny electromagnetic jammers and cyberweapons capable of interfering with the enemy's communication networks and targeting sensors (Rashid et al., 2023).

2.1.6 Information Operations

In the last decades, information operations have evolved and increased enormously in the context of hybrid warfare. They consist in the attempt by a state or a party to influence another state or specific populations by proposing purposeful narratives, through social media or other communication channels, to achieve a strategic objective (Whyte et. al., 2020). Although these

types of operations may appear to be low-intensity, they are actually high-intensity and can greatly affect the stability of a state. China and Russia are among the states that use these means the most in a systematic and aggressive manner—while other states traditionally used information operations more defensively. For instance, China's information operations involve exploiting the interests of social groups to achieve its objectives, particularly within the Western alliance system. These operations include the launch of campaigns involving inauthentic posts aimed at creating the perception of widespread support for a particular policy or narrative, as well as the promotion of specific narratives that benefit China's long-term political and economic objectives (NATO Parliamentary Assembly, 2025; Hunter et al., 2024). Russia is also well known for its use of information operations, particularly to influence electoral campaigns. This was famously seen in the Moldovan case but also in the 2016 US presidential election. Russia uses these operations to divide and polarize societies by exploiting pre-existing socio-political fault lines in Western societies. (Hunter et al., 2024).

In the realm of information operations AI is intensively used, both in offensive and defensive ways, enhancing the effectiveness and speed of such operations. AI produces realistic photos, videos, and audios, usually involving key figures, thereby creating more realistic false contents that can influence population's perception. Generative adversarial networks, a branch of AI algorithms, are used to produce realistic-looking data, disseminate it and tracking its performance. In operations involving trolls and bots, AI is used to spread false information—for instance through the creation of false accounts—often in combine with cyber activities like Distributed Denial-of-Service (DDoS) attacks (Hunter et al., 2024; Sayler, 2020). DDoS attack is a type of cyberattack that involves flooding a target server, service, or network with a massive volume of internet traffic from multiple sources, often compromised computers or "bots." This flood of traffic exhausts the target's resources, such as bandwidth, processing power, and memory, thereby rendering the service unavailable to legitimate users and disrupting normal operations (Holdsworth & Kosinski, n.d.). For defensive purposes, AI is used to monitor suspicious network activity, identifying attempts of influence and prevent the spread of disinformation. AI algorithms are also used to recognize manipulated image, videos and other type of false content (Sayler, 2020).

The effects of such operations can be extremely wide-ranging. Dumbacher (2025) highlights the risks associated with deepfake content in exacerbating threats to state stability. In an extreme scenario, deepfakes, disinformation, or the fabrication of false alerts related to military

mobilization by an adversary could increase the risk of escalation. Deepfakes have already been observed in ongoing conflicts. Shortly after Russia's invasion of Ukraine in 2022, a widely circulated deepfake showed Ukrainian President Zelensky urging Ukrainians to surrender. Similarly, in 2023, a deepfake caused viewers to mistakenly believe that Russian President Vladimir Putin had interrupted state television to declare a total mobilization (Dumbacher, 2025). Considering the significant social impact of these threats, the development of effective mitigation tools—potentially leveraging AI—is therefore essential.

Similarly to cybersecurity, AI-enabled information operations illustrate the evolution of non-conventional forms of destabilization. While such practices were already present in earlier hybrid conflict strategies, AI significantly enhances their scale, speed, and sophistication, thereby altering their strategic impact. Through automated content generation, micro-targeting, and data-driven influence campaigns, AI reshapes the informational domain into a more pervasive and difficult-to-regulate environment of contestation. This transformation calls for a fundamental rethinking of security. Addressing AI-driven information operations calls for a proactive and continuous effort to strengthen the broader security architecture, involving not only state institutions but society as a whole. Building resilience against disinformation and manipulation entails enhancing digital literacy and cross-sectoral coordination, thereby fostering a more robust and anticipatory security posture.

2.1.7 Logistics

Logistics is one of the fields in which AI has been implemented more easily, as it is a sector with a vast amount of data, involving manual and often repetitive tasks. Applications of AI in military logistics range from the management and optimization of supply chains to predictive maintenance, resulting in significant cost savings and improvements in operational readiness.

In predictive maintenance, AI is used to analyze historical data in order to predict future logistical needs and allowing for proactive resource allocation. It is used to monitor equipment performance and condition during operation, reducing failures and improving availability. For instance, the U.S. Air Force utilizes the ALIS system, developed by Lockheed Martin, to monitor its F-35 aircraft. The systems gather real-time information from the sensors installed on the aircraft, analyzes the data and advise maintenance technicians on whether it is needed to inspect and replace

specific components (Shin et al., 2019). The result is more accurate preventive maintenance and, consequently, reduced costs and unforeseen damages. Beyond F-35 aircraft maintenance, the U.S. military has actively adopted AI for smart maintenance concepts in several areas.

AI is also exploited for the automation of supply chain management and demand forecasting. Through its analysis capabilities, AI can automate various aspects—such as forecasting demand and managing inventory—making the process less prone to human error and more efficient. For instance, AI models are used to calculate the most optimized supply plans to save time and reduce costs (Rashid et al., 2023; Shin et al., 2019). Furthermore, AI and ML improve decision-making in logistical operations. AI systems can support military leaders by providing informed logistical decisions through real-time data analysis, enabling the selection of the most efficient logistics routes in terms of both time and cost (Cecchini, 2023).

In the field of logistics and equipment predictive maintenance, AI represents one of the most stable and comparatively low-risk areas of implementation. AI systems can enhance the quality, efficiency, and speed of logistical operations and material support, thereby contributing to improved operational readiness, reduced downtime, and optimized resource allocation. Compared to applications directly involved in targeting or real-time decision-making, the associated risks in this domain are relatively lower.

2.1.8 Training and Simulations

AI is utilized in military training and simulations to increase realism, personalize instruction, assist with planning and tactical development, and evaluate performance through game-based simulations and models.

AI is leveraged to train human personnel in a personalized manner, adjusting the difficulty level to match and challenge their competence level. It is used to create virtual simulations and improve real-life exercises, thereby improving the preparation of individuals and groups (Rashid et al., 2023). An example is the Virtual Training & Environments, an office of the U.S. Naval Research program, which creates virtual trainers for military operations in urbanized terrain. In the exercise, U.S. Marines are stationed in a simulated urban area and tasked with clearing a building with inside enemy soldiers. These virtual opponents test the trainees (Rashid et al., 2023). AI is also used to create training programs that adapt to the individual, tracking his learning progress and offering tailored activities. ML is used to create individual profiles that learn and adapt over time.

This concept is referred to as hyper-personalization (Rashid et al., 2023). Therefore, by enabling adaptive learning environments and realistic scenario modeling, AI supports better overall preparedness and personnel readiness.

Healthcare on the battlefield

AI is also used to assess medical data from soldiers and monitor their health. AI can monitor physiological and biomedical parameters, such as body temperature, heart rate and EEG, in real-time through sensors located close to the individual body. In this way, health experts can continuously access soldiers' health status and provide real-time responses. Moreover, AI systems are used to identify threats, such as approaching bullets, bombs or damaging waves, and alert the individual before they can do so themselves. This is possible by the integration of smart helmet, smart uniform and smart eyewear sensors (Rashid et al., 2023).

2.2 The Debate over the Impact of Military AI

While efforts are underway to increasingly implement and integrate AI technologies into the military domain, the extent of such a revolution is debated, and exaggerations often obscure the real capabilities of AI. Rickli and Mantellassi (2023) propose a useful distinction by dividing the researchers into three school of thought: enthusiasm, denial, and pragmatism. The debate presented contributes to a more nuanced understanding of the implications of AI integration in military contexts. Indeed, these three schools do not differ on the recognition of the recent development and implementation of AI technologies in the military field. Rather whether and to what extent military AI will impact the character of conflcits by offering decisive advantages to their adopters.

Enthusiasts

Among the enthusiasts are scholars who believe that AI will lead to a revolution in the character of warfare, from the smallest to the largest conflict, by changing the way states fight and transforming the full spectrum of military activities (Rickli & Mantellassi, 2023). In their view, AI has military applications that confer decisive advantages to those who adopt it.

First, AI augments the robotization of warfare by increasing the deployment of autonomous systems. According to enthusiasts, autonomous vehicles and weapons allow forces to conduct sophisticated battlefield tactics and operations, rapidly exploiting emerging opportunities (Davis,

2019). Because autonomous systems are able to react to shifting events faster than is humanly possible, operations would be conducted at machine speed, accelerating the entire process and response times (Rickli & Mantellassi, 2023). AI-enabled systems could allow more precise and effective operations, therefore reducing costs. An example is provided by drone swarms. Furthermore, enthusiasts argue that advances in AI and machine autonomy will enable these systems to act as one on the battlefield, in a more coherent and effective manner than humans possibly could.

In fact, in the war between Ukraine and Russia, drones have surpassed conventional arms and now dominate the battlefield. According to Ukrainian commanders, drones have caused more casualties and destroyed more armored vehicles than all traditional weapons of war combined—including sniper rifles, tanks, howitzers, and mortars (Santora et al., 2025). However, at present, the majority are basic models with limited use of AI. Kateryna Bondar, speaking at a CSIS event (Center for Strategic and International Studies, 2025), explained that the most common drone used in this conflict is the FPV drone (first-person-view drone). In its basic form, it consists of a quadcopter equipped with a remote-control station that allows the operator to see what the drone sees during flight. Nevertheless, AI-enabled drones are also beginning to spread, vehicles capable of autonomous navigation and automatic target recognition. Although AI implementation is still limited, both Ukraine and Russia are actively working to expand their deployment. Therefore, if even relatively simple drone technologies have demonstrated the capacity to generate strategic advantages and reshape the character of conflict, the prospective deployment of AI-enabled systems is likely to exert even more profound effects on the conduct of warfare, potentially introducing new forms of conflict characterized by enhanced precision, adaptability, and decision-making at machine speed. It is therefore plausible that AI will significantly alter the speed and character of battlefield operations.

Moreover, according to enthusiasts the nature of war's brutality could also evolve. Mallick and Gen (2019) argue that if machines become predominant because of their lethality and accuracy, making ordinary soldiers obsolete, then war could become a case of "machines being violent to other machines". Precision in attacks could increase, limiting damages to civilians, non-combatant and humans in general. Simultaneously, AI may alter war's nature also by increasing the use of hybrid tactics—such as disinformation campaigns, deepfakes, and cyberattacks. This could lead to more sophisticated and less visible forms of warfare, in which covert means are employed to

generate internal disorder and dissent—such as through psychologically manipulative forms of conflict—thereby undermining a state from within.

Actually, the rise of hybrid threats is an immediate and pressing reality. Non-conventional forms of destabilization are increasingly becoming significant sources of security challenges, calling for a substantial shift in strategic perspective. In this regard, the traditional reactive approach characteristic of the NATO member states appears inadequate against these types of operations, which instead require a preventive and continuous posture that actively counters destabilization efforts aimed at eroding a state or institution from within (European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE, n.d.).

Furthermore, enthusiasts emphasize that the capacity of AI to analyze vast quantities of data at speeds unreachable for humans holds important military potential (Rickli & Mantellassi, 2023). This, as we have seen, is one of the most researched aspects. The ability to navigate battlefield data and provide timely analysis is crucial to modern militaries. With increasing data flows—from the rise in the number of machines deployed, but also from sources such as the internet—AI can accelerate the analyses and processing of information, providing more comprehensive assessments. Strategic technological application could help contain two intrinsic warfare elements: the "fog of war", a term used to indicate uncertainty and confusion, and its "friction", the unforeseen difficulties that make simple tasks complex (Rickli & Mantellassi, 2023). In this way, situational awareness is enhanced and the decision-making process facilitated.

Finally, according to enthusiasts, the adoption of AI in the military will have critical consequences for the global balance of power. This is already becoming apparent. The U.S. and China are engaged in a race for AI dominance, driven by the conviction that AI provides a strategic advantage.

Deniers

Deniers sustain that AI will not change the nature of war and will have little impact on this character. Technical and organizational difficulties limit the potential of AI on the battlefield (Rickli & Mantellassi, 2023). AI technologies are still too immature and far from represent decisive assets in the battlefield: they struggle in real-world context characterized by unpredictability and uncontested environments. In the military realm, still do not exist large datasets needed to train ML algorithms. Svenmarck et al. (2018), presenting their work on the opportunities and limitations of applying AI in military contexts during a NATO conference, highlighted that there are not enough

“real world, high quality, sufficiently large datasets” that can be used to train algorithms and develop successful ML-based military AI applications.

This statement is quite accurate. For instance, the Maven system’s accuracy (see Section 2.1.1) is highly dependent on the quality and quantity of its training data and, in the Middle Eastern desert, the AI model reportedly misidentified trucks as trees or valleys (Dina, 2024). Moreover, certain AI systems are vulnerable to adversarial inputs and can be easily fooled with relatively easy techniques—for instance, by modifying one pixel to completely misdirect image recognition software (Rickli & Mantellassi, 2023). Thus, the availability of adequate and high-quality data for training AI systems represents a noteworthy challenge.

Deniers view the same Ukrainian drone deployment as evidence of AI’s immaturity and the limits of current technology. They point out that the drones used in the war between Russia and Ukraine are still very basic models—mainly piloted by humans, minimally networked, and small in size, with mostly localized effects. Even those equipped with AI and ML still require significant human intervention. Both sides are engaged in an innovation–emulation cycle. Since many of these technologies are commercial or dual-use and thus easily accessible, once one side deploys them militarily, the other can quickly emulate and implement them as well (Pettyjohn, 2024). Deniers point out the fact that the integration of AI remains a highly expensive and time-consuming process. Developing military-grade AI capabilities requires the creation of new and complex algorithms, significant computing power, vast data storage infrastructure, and extensive long-term testing to train AI systems to operate and learn from diverse battlefield environments (Stepanenko, 2025). As a result, deniers emphasize the considerable challenges of effectively implementing AI in military operations and highlights the continued reliance on human oversight and conventional strategies, underscoring claims of an AI-driven military revolution.

These are valid considerations, as AI is highly data-, capital- and energy-intensive; however, rather than preventing its development altogether, the greater risk is that only a small number of states will possess the capacity to mobilize the resources required, potentially exacerbating existing power asymmetries.

Rickli & Mantellassi (2023) emphasize that beyond technical limitations, the success of AI in the military realm depends on the organizational difficulties in integrating autonomy in military concepts and doctrine. A research from Borchert et al. (2021), by studying recent conflicts in Libya, Ukraine, Syria, and Nagorno-Karabach, points out that the extent to which autonomous systems can give an advantage depends on how they are integrated. According to them, the technology

needs to be embedded in the broader cultural, conceptual, and organizational context. Therefore, innovation alone it is not sufficient to transform warfare.

Furthermore, AI-enabled military presents an issue of transparency and predictability given by its nature of "black box". As already explained, at the current state we are unable to explain the reasoning which leads the system to a particular outcome. This can present several challenges once applied in the military field, where explicability and predictivity are essential to the commander's decision making, therefore limiting and slowing down its application (Rickli & Mantellassi, 2023). Commanders may hesitate to rely on systems of which cannot explain the output, particularly when they are held accountable for the resulting actions.

This is a highly valid consideration, as the opacity of AI systems constitutes a significant limitation of these technologies. Nevertheless, rather than halting or significantly constraining their development, it is more likely that states and military institutions will adopt adaptive strategies to integrate such systems while mitigating their opacity.

Finally, according to this school of thought, AI might increase "fog of war" rather than decrease, by increasing the amount of information, with events going too fast for human to follow, therefore increasing battlefield unawareness (Rickli & Mantellassi, 2023).

Pragmatics

The perspective of the pragmatics holds that AI can play a significant role in the military domain by facilitating actions and operations (Rickli & Mantellassi, 2023). AI's capabilities in data collection, analysis, and prediction can be exploited in intelligence gathering and operational planning—particularly at the tactical levels, and especially in controlled environments. Nevertheless, AI shows limitations when facing unpredictable and rapidly changing conditions, and it displays a relative lack of resilience against adversarial attacks. For these reasons, pragmatists argue that the areas where AI can be most successful are those operating in uncontested domains—such as communication, training, resource allocation, and logistics (Maxwell, 2020). AI can lead to predictive logistics and autonomous convoys, enable more efficient repairs, enhance time and resource allocation, provide more personalized and realistic training for military personnel, and streamline the processing of data from multiple sources. However, its application outside these domains appears more complex and tortuous (Rickli & Mantellassi, 2023).

Additionally, pragmatists emphasize that the implementation of AI in the military realm also depends on political and societal debates. Nations and the international community remain

highly divided over the future of AI in warfare, particularly regarding LAWS. Many countries, such as Austria, Brazil, Chile, Morocco, Egypt, and Colombia, have called for a ban on LAWS. The UN Secretary-General António Guterres also has repeatedly called for a global ban. Other states opposed to the development of banning regulations, among these the U.S., the U.K., Australia, Russia, India, Israel, Japan, South Korea, and North Korea. They argue that applying AI to weapons can offer significant strategic advantage and will not further war but could, instead, reduce collateral damage to non-combatants and civilians (Autonomous Weapons, 2025; Rickli & Mantellassi, 2023). Pragmatists emphasize that institutional factors can significantly constrain the implementation of AI.

Furthermore, they stress that advancements in AI go hand-in-hand with innovations in other technologies—convergence with these technologies may lead to outcomes we cannot yet predict.

Pragmatists also acknowledge the black box dilemma but highlight the ongoing research in XAI which is working to increase transparency. Also, they argue the issue can be bypassed by limiting AI deployment to domains do not require explainability, such as logistic data fusion or predictive maintenance.

Finally, many researchers underscore that even if AI assumes a dominant role in warfare, strategic development will remain primarily human-driven. Humans will retain command, and AI is better framed as a "strategic counsellor". AI can provide important advantages to adopters by anticipating and identifying risks associated with strategic options, recognizing patterns that humans might overlook, and responding or adapting more rapidly (Rickli & Mantellassi, 2023).

2.3 Challenges and Ethical Dilemmas of Military AI

The debate presented already introduced how the full implementation of AI in the military domain entails not only operational advantages but also a range of significant challenges. These challenges emerge at both the technical–operational level and within the broader ethical, legal, and governance debate. Given the direct implications of military applications for states security, human life, escalation dynamics, and international stability, these issues acquire particular relevance.

Technical and operational challenges

AI development is based on key enabling resources, including data, computing power, and energy. Therefore, AI development highly depends on these resources and, with AI becoming increasingly strategically important for security purposes, external dependency on such resources

can represent a significant vulnerability and constraints on states' autonomy in the development, deployment, and sustainment of AI capabilities particularly for defensive purposes.

First, high-performing AI systems require highly specialized talent and expertise capable of designing, training, and maintaining advanced algorithms. At present, despite the growing research capability within defense institutions and the establishment of dedicated operational units in many states, the most advanced AI models remain largely concentrated in the private technology sector and in only a small number of states—most notably the U.S (Clapp, 2025*b*). Whereas many twentieth-century breakthroughs originated in military R&D before diffusing into civilian use—such as the Internet—contemporary AI innovation is largely driven by the commercial sector. This shift has reinforced dependence on private actors for the design and development of military AI systems. For example, the U.S. system Raven Sentry used in Afghanistan in 2020 (see Section 2.1.1), involved the recruitment of experts from Silicon Valley (The Economist, 2024). This growing dependence on private actors for the design and development of military AI systems raises significant concerns related to security, accountability, reliability, and state control over critical defense technologies. Private actors may hold interests and preferences that diverge from those of the governments implementing such technologies. Their decisions could influence the dynamics of conflicts. For example, such as in the case of Elon Musk's Starlink satellites, whose service restrictions in 2022 during the war in Ukraine reportedly influenced military capabilities (Roulette et al., 2025; Clapp, 2025*b*). These dynamics call into question the practicability of regulating these technologies through traditional arms control regimes (Lonergan, 2025) and raises serious concerns regarding the influence of private companies over national security decision-making and the extent of state control over critical military infrastructures (Clapp, 2025*b*).

Second, the hardware components of military AI systems—data centers, semiconductors, and advanced microchips—introduces additional strategic vulnerabilities. AI development depends on critical raw materials, including gallium, germanium, and palladium. China currently controls large portion of the global market of these materials. As a result, many countries—including the U.S. and the European states—remain highly dependent on imports of these materials. Following the U.S. export controls on advanced semiconductors aimed at restricting China's access to cutting-edge technologies, China responded by imposing restrictions on the export of gallium and germanium (Baskaran & Schwartz, 2024; Allen & Goldston, 2025). This dynamic underscores how competition over AI increasingly extends beyond software and algorithms to encompass material

supply chains. If AI is going to play an increase role in national security, the risk is of dangerous dependences on critical material that directly affects states' capacity to develop and sustain military AI capabilities. A dependency that translates into strategic and security risks.

Third, robustness and reliability issues of AI systems must be considered. At the time of writing, some studies suggest a tendency for advanced AI models to develop emerging and unforeseen behaviors, including optimization strategies that violate predefined constraints, unexpected interactions among interconnected AI systems, and forms of goal mis-generalization not anticipated during training (Archivio Disarmo (IRIAD), 2025)[7]. In the military domain, AI systems appear to be particularly susceptible to compromises of system integrity (Farrar, 2025). Manipulated, corrupted, or biased training data, as well as the limited diversity of military-specific datasets, can lead to erroneous predictions or decisions with potentially catastrophic consequences. With regard to this possibility, the U.S. Department of Defense's Defense Innovation Board (2019) underlines that "AI systems should have an explicit, well-defined domain of use, and the safety, security, and robustness of such systems should be tested and assured across their entire life cycle within that domain of use". Similarly, the Organisation for Economic Co-operation and Development (OECD) stresses that "AI systems must function in a robust, secure and safe way throughout their life cycles and potential risks should be continually assessed and managed" (OECD, n.d.). Therefore, rigorous testing to assess the adequacy and efficiency of AI systems across various operational contexts is required to ensure that they meet the demands of real-life military applications.

Ethical and Governance Dilemmas

Alongside technical constraints, the rapid development of AI systems for military use, together with the potentially disruptive capabilities of these technologies—first and foremost those stemming from AWS and LAWS—is raising also significant ethical concerns. In the military domain, technological development is reshaping traditional procedures and operational approaches. Contemporary battlefields, such as the conflict in Ukraine, have increasingly

[7] The report presented by Archivio Disarmo (IRIAD), *Lo stato dell'Intelligenza Artificiale in ambito militare e le prospettive di regolazione a livello nazionale, europeo e internazionale* (2025), specifies that research in the area is still ongoing, and it is possible that some of these effects may diminish, fail to be confirmed, or more fully explained though more in-depth analyses. Nevertheless, the possibility of such effects calls for considerable attention.

functioned as testing grounds for emerging technologies, accelerating experimentation and deployment in ways that challenge established operational doctrines and question ethical norms. Added to this are the tensions arising from the practical implementation of certain ethical principles. For instance, transparency and traceability emerge as key ethical requirements that nevertheless need to be balanced against the secrecy constraints necessary to protect sensitive information and strategic advantages. This necessitates the development of appropriate mechanisms that ensure accountability and compliance with legal standards, while protecting sensitive information relating to military operations. (Taddeo, 2024; European Defence Agency, 2025). In recent years awareness of the need for a structured ethical discourse on military AI has grown significantly.

A primary significant issue concerns accountability and responsibility. Determining responsibility for decisions made by AI-enabled systems is complex, yet essential as such systems increasingly assume prominent roles within military processes. Responsibility cannot be attributed to machines and human agents should be held accountable for the development, testing, use, and behavior of any AI military system (NSCAI, 2021). Accountability must also be differentiated across the AI lifecycle, including developers, military authorities responsible for deployment, and operators in the field. If an AI system causes unintended harm, determining whether the responsibility lies with the model developer, the authority that decided to deploy the system within the armed forces, or the individual who operated, becomes necessary (Clapp, 2025*b*). In order to minimize accountability gaps, major institutions emphasize the necessity of maintaining meaningful human judgment throughout the entire AI life cycle. The DoD's Defense Innovation Board of 2019, in its document *AI principles*, identifies three levels of responsibility. The first level comprises the individuals involved in the development, design, acquisition, testing and evaluation of AI systems. The second level concerns the C2 structures that use AI system in the conduct of hostilities. The third level relates to post-hostilities redress mechanisms, where responsibility ultimately rests on the exercise of "appropriate human judgment". Similarly, also NATO emphasizes the importance of maintaining meaningful human judgment in the use of AI by articulating the principle of *Responsibility and Accountability*, stating that "AI applications will be developed and used with appropriate levels of judgment and care; clear human responsibility shall apply in order to ensure accountability" (NATO, 2024). Therefore, human involvement throughout the entire AI lifecycle is essential to mitigate operational risks and address accountability

challenges, provided that human involvement is meaningful and that operators are adequately trained to recognize errors or biases in AI systems. Indeed, a present risk lies in overreliance on the system, which may lead to reduced vigilance and diminished critical oversight. Human operators must therefore be properly prepared and aware of all possible constrains of the system.

Closely related is the ethical concern related to the lack of transparency of AI and its "black box" nature. Indeed, it is often not possible to fully understand how an AI system processes information, making it difficult to anticipate its behavior in specific deployment contexts or to examine outcomes when they are unintended or undesirable (Taddeo, 2024). This opacity it implies reliance on machine-generated outputs without a clear understanding of how a particular result was produced. In military contexts—where decisions may involve escalation dynamics or civilian harm—delegating authority to systems lacking transparency raises serious ethical and legal concerns. Moreover, crucial underlying issues connected to the lack of transparency are the risk of bias, distortion, and adversarial manipulation, as AI systems may reflect design choices, unrepresentative or skewed training data, or even adversarial interference with the system.

This opacity poses noteworthy security challenges on the control of the systems and on their adoption in the military domain. Appropriate responses to these challenges include the introduction of adequate safeguards and control mechanisms. Major institutions—including the U.S. DoD, NATO and the European Defence Agency—advocate for sustained and informed human involvement, ensured in both the deployment and use of such systems (Defence Innovation Board, 2019; NATO, 2024; European Defence Agency, 2025). Thus, also in this case, meaningful human presence emerges as a key element in response to ethical challenges. Human involvement must be meaningful and informed, grounded in a clear understanding of system capabilities and limitations. In addition to this, a risk-based approach to deployment is essential. An example in this regard is the framework proposed by the EU AI Act (which will be discussed in Section 4.4). The use of AI systems should be restricted in particularly sensitive and high-risk contexts, while their reliability strengthened through continuous testing, validation, and training, until sufficient levels of transparency, control, and trustworthiness are achieved. Such an approach helps prevent potential harms in sensitive domains while allowing for the gradual and responsible advancement of the technology.

A further ethical and legal concern relates to compliance with IHL. Military AI systems must adhere to fundamental principles such as proportionality, according to which military actions must be proportionate to the threat, and distinction, which requires that civilians be clearly differentiated from combatants (Clapp, 2025*b*). In this regard, the United Nations General Assembly adopted Resolution 79/L.77 on AWS in 2024, highlighting the risks associated with a potential arms race, as well as the ethical and humanitarian implications linked to the absence of meaningful human control in military decision-making. The resolution calls for strengthened international oversight of AWS and for ensuring compliance with IHL. It was adopted with 166 states voting in favor, 3 against (Belarus, North Korea, and Russia), and 15 abstentions (Saudi Arabia, China, Estonia, Fiji, India, Iran, Israel, Latvia, Lithuania, Nicaragua, Poland, Romania, Syria, Türkiye, and Ukraine) (United Nations General Assembly, 2024). However, the resolution is non-binding and lacks enforcement mechanisms, limiting its practical impact. The rapid development of AI technologies is deeply dividing states, with Europe and many countries of the Global South advocating for a more ethical approach, while major technological powers remain reluctant to accept binding legal constraints. These fragmented and unilateral regulatory efforts are insufficient to address the challenges posed by military AI—particularly LAWS—and coherent international standards and guidelines are required to regulate their development and use, and to ensure that such technologies are rigorously tested for full compliance with IHL.

Finally, AI raises broader considerations regarding strategic stability. AI technologies can streamline decision-making processes and significantly reduce the time required to reach operational decisions. While this acceleration can represent a substantial asset, it also introduces new risks. In particular, the compression of decision-making timelines may facilitate escalation dynamics by limiting opportunities for decompression, like diplomatic de-escalation. AI-processing rapidity may lead to rapid and increasingly automated response cycles, allowing crises to intensify faster than human intervention can effectively contain them (Archivio Disarmo (IRIAD), 2025). Such dynamics are especially concerning in high-stakes domains, most notably in relation to nuclear weapons. To mitigate these dangers and maintain meaningful control, safeguards have been introduced to ensure that humans retain final decision-making authority over the use of force. The 2022 U.S. National Defense Strategy explicitly states that a human must remain in the loop for any decision to employ or terminate the use of nuclear weapons. Similarly, former U.S. President Joe Biden and Chinese leader Xi Jinping affirmed in parallel statements that “there should

be human control over the decision to use nuclear weapons" (Renshaw & Hunnicutt, 2024; Dumbacher, 2025).

3. AI in the Healthcare Sector

Just as in the military, AI is proving particularly valuable in the healthcare sector, helping to overcome the numerous challenges it has faced in recent years. Indeed, the healthcare sector is currently undergoing a significant period of revolution and transformation. This shift is primarily driven by challenging global factors, including increased demand for healthcare services, ever-rising total healthcare spending, and an escalating shortage of professionals in the sector.

The demand placed on healthcare systems has increased significantly in recent decades, with the rise in population life expectancy being one of the main drivers. Over the past century, average life expectancy at birth has increased from less than 50 years to an average of 80.8 years in EU Member States, with some countries reaching an average of 83 years. The proportion of the population aged 65 and over has risen from 16% in 2000 to over 21% in 2023, with projections indicating a further surge to nearly 30% by 2050 (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). The growth of life expectancy is resulting in increasing incidence of chronic and complex conditions. According to the OECD and the European Commission (2024), 40% of EU citizens aged 65 and above live with at least two chronic conditions. An aging population and the elevated presence of chronic conditions translates into an increased demand for healthcare services. In such a scenario, the healthcare system must be organized in terms of its structures and workforce in order to cope with this higher demand. While this increased demand on healthcare systems has led to higher healthcare expenditure, which has become one of the largest government expenses, shortages in the healthcare workforce are also present. Healthcare expenditure has notably increased in most European countries in recent decades, accounting for 8.1% of gross domestic product, whereas in the U.S. it amounts to 18.3% (OECD & European Commission, 2024). Nevertheless, the shortage of healthcare workers poses a serious limit to the capacity of the healthcare sector. Several European countries reported shortages of doctors in 2022 and 2023 (OECD & European Commission, 2024). This significantly impacts the ability of healthcare structures to cope with higher demands on the sector. Shortages of healthcare professionals (HCPs) increase pressure on healthcare systems, leading to high levels of burnout among personnel, lower service quality, and longer waiting times for patients.

Emerging technologies—among which AI, ML and big data analytics—are slowly entering the healthcare sector and have the potential to transform it significantly, making it more efficient and smarter. AI has the potential to help address shortcomings in the healthcare sector and

contribute to achieving the goal of an increasingly predictive, preventive, personalized, and participatory healthcare—the so-called 4P model. By improving the quantity and quality of processed data, these technologies can enhance operational efficiency and alleviate the pressure on healthcare systems. Integrated into the administrative logistics sector of a hospital, these technologies can handle all routine operations, such as bed allocation and updating medical records. Their implementation in healthcare also seems to have great potential for improving the quality of healthcare services, from prevention to surgical operations.

The features used are similar to those used in the military field. First, AI's capacity to analyze vast amounts of health data within a relatively short timeframe can significantly accelerate medical analysis and data management processes. This feature can be applied across several areas, including diagnostics and prevention, medical research and—the most immediate application—administrative tasks. AI-driven data analysis can support staff rota planning and hospital bed allocation through more efficient analysis of healthcare infrastructure and patient flow, as well as contribute to scientific research, particularly in the discovery of new drugs and vaccines. Second, emerging technologies' capability of identifying patterns or anomalies proves particularly valuable, especially for disease detection and the prediction of health trends. When applied to medical imaging, this capability enables the faster identification of abnormalities, such as cancer or other diseases, thereby facilitating early diagnosis and timely intervention while conditions are still at an early stage. Moreover, the use of AI reduces task-related fatigue and enhances consistency in activities that are especially prone to human error. Finally, Chatbots and NLP technologies further enhance telemedicine by assisting patients in home settings and streamlining remote doctor–patient interactions, contributing to a more personalized and accessible healthcare system.

3.1 Applications of AI in Healthcare

AI implementation spans numerous domains, including surgery, medical imaging and diagnostics, virtual patient care, treatment design, administration and workflow management, and medical research and drug discovery. Figure 2 shows the possible areas of implementation of AI systems, according to hospitals, HCPs, and developers. The area that seems to have the most natural implementation is the streamlining of administrative tasks and work optimization, followed by diagnostics. Nevertheless, the figure shows that developers are optimistic about implementing AI systems also in areas such as treatment prediction and personalized treatments. The fundamental goal of AI in this context is to augment the capabilities of human professionals, with the intent of achieving cost reductions and providing better healthcare outcomes. Also in this case, AI is seen as a support instrument that can enhance the work of human operators by making it faster and more accurate, though it cannot replace them. The human and personal aspects of medicine are important, and it is crucial that these aspects are not lost.

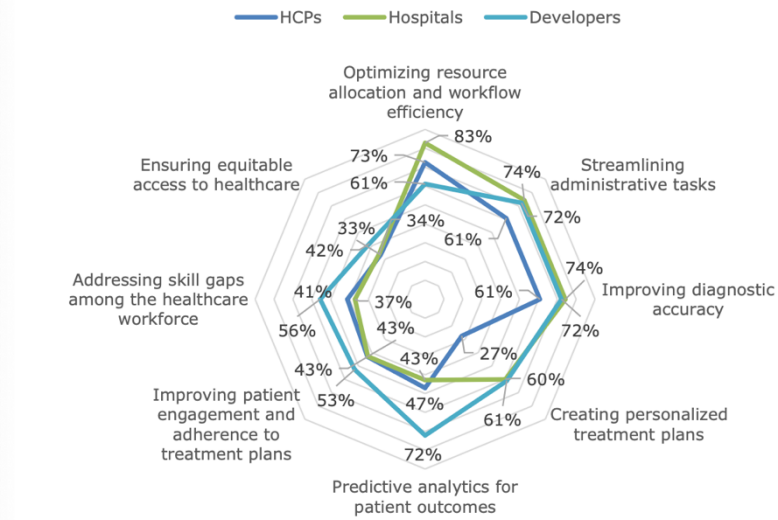

Figure 2 – Healthcare needs that can already be addressed by existing AI solutions according to HCPs, hospitals, and AI developers (source: European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025)

3.1.1 Administrative Assistance

One of the most straightforward areas for implementing AI is the management of administrative tasks in hospitals and healthcare centers. The administrative workload in hospitals has increased significantly in recent years, driven by the growing volume of hospital activities and expanding bureaucratic requirements. At the same time, hospitals have become increasingly digitalized and paperless. However, digitalization has not necessarily resulted in administrative simplification. Indeed, although such transformation has brought many benefits, it also presents some challenges. A number of studies reveal that a significant proportion of HCPs' working time

is spent on non-medical activities, such as administrative tasks and logistical duties. A study conducted in the US involving 200,081 HCPs from 396 organizations revealed that HCPs spend an average of 5.8 hours out of an 8 hour shift actively working on the electronic health records (EHRs), meaning the EHRs have actually resulted in an increase in the administrative burden for HCPs globally (Holmgren et al., 2024).

AI can prove particularly useful in reducing administrative burden, in this way allowing clinicians to focus on more important activities such as direct patient care. This is also one of the simplest and least risky uses. In the logistical management of healthcare facilities, AI models can be employed to support operations such as back-office functions, patient scheduling, hospital bed allocation, insurance claim verification, and staff shift management. These technologies can be leveraged to analyze large volumes of data and identify patterns in hospital patient flows, enabling improvements in the efficiency of logistical processes. For instance, managing bed availability is a critical aspect of hospital administration, as it directly affects the quality of patient care, staff workload, and overall healthcare operational efficiency. Hospitals are often overcrowded, particularly in the aftermath of the Covid-19 pandemic. In this regard, AI technologies can provide valuable support by predicting bed availability and suggesting the prioritization of patients with more serious needs, as well as the more efficient allocation of healthcare resources. The Johns Hopkins University Hospital stated that implementing AI technology allows them to assign emergency room beds 30% faster, reduce transfer delays from operating rooms by 70%, enable ambulances to pick up patients from other hospitals 63 minutes earlier, and improve their ability to accept patients with complex medical conditions from regional and national hospitals of 60% (Mudgal et al., 2022). Test with Microsoft's Cortana digital assistant used advanced analytics and predictive technologies to identify patients at risk in intensive care. It was able to monitor 100 beds across six intensive care units. (Mudgal et al., 2022).

In staff shift planning, AI systems can support more efficient scheduling. Indeed, planning staff rotas represents a critical challenge in the healthcare sector due to understaffing, frequent modifications to assigned schedules to accommodate the inherent unpredictability of patient care, personal preferences, and unexpected last-minute staffing demands—dynamics that became especially evident during and after the Covid-19 pandemic. In this context, the analytical capabilities of AI can substantially enhance scheduling performance by enabling hospitals to more effectively balance staff availability with patient needs. In critical care environments, AI-based

systems are used to anticipate practitioners' workloads in intensive care units and emergency departments, ensuring that appropriately skilled personnel are allocated to specific shifts (Liu et al., 2018). Such systems can incorporate multiple variables, including healthcare workers' competencies, specific skill sets, as well as factors related to staff satisfaction and fatigue.

AI can also be applied through predictive maintenance models to monitor medical equipment. Indeed, an average hospital hosts between 5,000 and more than 10,000 different types of medical equipment—such as MRI machines, CT scanners, X-ray machines, and ventilators—which are essential for diagnosing and treating patients (Malik & Solaiman, 2024). These devices are highly costly and require regular maintenance throughout their lifecycle, representing a significant additional operational expense. For this reason, hospitals and medical device manufacturers are increasingly embracing predictive maintenance approaches enabled by AI technologies. These systems leverage data already collected from medical equipment to identify the progression patterns of damage and to predict the future condition of devices. For instance, in a hospital in the United Arab Emirates, predictive maintenance models have been used to anticipate potential issues with the Vitros Immunoassay Analyzer—a medical device used to perform blood tests for conditions such as HIV and hepatitis. Early detection of anomalies enables timely intervention, thereby minimizing downtime and ensuring continuity and reliability in diagnostic testing for patients (Malik & Solaiman, 2024).

Such enhancements in hospital's logistics and administrative management translate into greater operational efficiency and a reduced risk of systemic collapse under pressure, thereby strengthening the capacity to respond to health security challenges. In the event of a health crisis, more efficient coordination of resources, personnel, and infrastructure enables healthcare systems to respond more effectively without becoming overwhelmed.

Beyond logistical functions, AI technologies can support healthcare professionals in other daily non-clinical tasks, such as documenting patient encounters and updating medical records. These systems can assist HCPs in clinical and administrative workflows by automatically extracting and structuring information from therapeutic notes, identifying relevant data from previous medical records, and collecting documented patient encounters (Al Kuwaiti et al., 2023). In turn, the accuracy of ML algorithms may benefit HER systems. A study conducted at The Permanente Medical Group, in the US, has shown how AI can offer useful support to HCPs by producing high-quality medical records. The documentation, transcriptions, and summaries of

encounters produced by the AI system have proven to be mostly of good quality, offering significant help to clinicians, although in some cases requiring revision. The implementation of AI systems has resulted in a significant reduction of the time spent by primary care physicians on clinical documentation outside of working hours and on notetaking during visits (Tierney et al., 2024; European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025).

3.1.2 Diagnostics and Prediction

AI is proving particularly useful in disease diagnosis, supporting a shift toward more predictive and preventive medicine by enabling early detection and reducing diagnostic errors. A delay in diagnosis, and consequently in treatment, can result in reduced treatment effectiveness and significant costs in therapeutic, economic—according to an OECD working paper by Luke Slawomirski et al. (2025), diagnostic errors financially account for the 17.5% of total healthcare expenditure—and human terms. Studies have shown that AI can improve the speed and accuracy of diagnosis in several medical specialties and appears as the field with the most promising implementation. In medical imaging, AI systems can analyze various sources—such as X-rays, CT scans, echocardiography, and mammography—and highlight the possible presence of diseases or anomalies, as well as detect minor patterns that humans would completely overlook. The result is faster and more accurate medical diagnosis (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025; Väänänen et al., 2021).

Because of their ability to detect anomalies in patterns, AI, ML, and DL hold enormous potential in digital pathology and radiology, which is considered one of the most mature fields for AI applications. This is due not only to the pattern recognition capabilities that are particularly well suited to the field, but also to the vast amounts of digital data accumulated over the years, facilitated by the widespread adoption of international standards in medical imaging—such as DICOM (Digital Imaging and Communications in Medicine) and systems like PACS (Picture Archiving and Communication Systems)—which therefore provide a large database on which these systems can be trained.

Multiple studies have confirmed the benefits in terms of time and quality in the implementation of AI in radiology analysis. A study at a German university concluded that AI tools used in in chest radiographs analyses reduced the time taken to report findings from 80 minutes to

35–50 minutes (Van Leeuwen et al., 2022). A study conducted by the National Consortium of Intelligent Medical Imaging in Oxford found that AI-assisted image analysis algorithm improved junior readers' proficiency in identifying pneumothoraxes on chest X-rays, achieving a level of accuracy comparable to senior or consultant readers (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). In the field of ocular diseases diagnostics, AI tools have also been developed for identifying signs of diabetic retinopathy, maculopathy, and exudates (Mudgal et al., 2022). For instance, the Moorefield´s Eye Hospital in London has developed AI solutions to detect ocular diseases by optical coherence. The results show a referral decision by the algorithm 94% accurate (Mudgal et al., 2022). Finally, Gudigar et al. (2021) found that several medical imaging tools using AI techniques—including X-ray, computed tomography, and ultrasound—have significantly contributed to countering the spread of Covid-19 by aiding in early diagnosis.

In addition to medical imaging, AI algorithms also can offer support in diagnoses by the analyses of health data, medical history, genetic information, lifestyle factors, and patient symptoms, supporting tailor prevention and treatment strategies, in this way improving outcomes. In the field of diabetes prevention and treatment, AI models can assist with the management by monitoring glucose levels. The Guardian Connect system by Medtronic[8] is the first AI-powered continuous glucose monitoring system of its kind. Based on ML predictive algorithms, the system can forecast significant changes in blood glucose 60 minutes before it occurs. Through a sensor placed on the abdomen, the sensor monitors blood sugar with an interval of five minutes. This enables proactive control of blood sugar levels, allowing patients to normalize the levels before a hypoglycemic event occurs (Väänänen et al., 2021).

NLP have been developed to interact with patients, offering support in understanding symptoms and addressing needs. In psychiatric diagnostics, the Columbia University's New York State Psychiatric Institute and the IBM Watson Research Center have developed an ML chatbot capable of detecting with 100% accuracy the development of psychosis in susceptible individuals by analyzing speech patterns. Traditional diagnosis achieves an accuracy rate of 79% (Väänänen et al., 2021)[9].

[8] Medtronic is an American–Irish medical device company. In recent years, it has been researching AI development for healthcare products and therapies to improve outcomes.

[9] It is important to note that this system was specifically designed for this purpose, unlike other AI tools that are sometimes incorrectly used for medical applications. Indeed, ChatGPT and similar large language models are increasingly used by the general public to seek medical advice, which has become a growing source of concern. Such models are not specifically designed, validated, or regulated for clinical diagnostic use. Consequently, reliance on these

Finally, AI proves valuable in the diagnostic of rare disease. A diagnosis of a rare disease––which affects approximately 400 million people worldwide (Mudgal et al., 2022)—can take up to five years. This process is often a time-consuming, costly and emotionally draining process for patients and their families (National Gaucher Foundation, n.d.). The algorithm designed by 3Billion in 2019 can diagnose rare DNA-based conditions and test for up to 7000 diseases simultaneously in suspected cases (Mudgal et al., 2022).

Cancer early detection and diagnosis

Among the various areas, one of particular interest for the benefits that the implementation of AI could bring is the early detection and preventive diagnosis of cancer. These technologies appear to be valuable tools for early detection, improving diagnostic accuracy, and personalizing treatment approaches. For these reasons, research in this field is rapidly expanding, and AI seems to play a significant role in improving treatment outcomes and reducing mortality.

Early detection in the case of cancer can be decisive for its treatment and therefore an increase in survival rates. The diagnosis involves various components depending on the type, location, and suspected stage of the tumor. Methods for detection include imaging techniques (e.g., x-rays, mammography, etc.), blood tests, endoscopic procedures, biopsies and physical examinations, among others. AI algorithms have proven effective at diagnosing certain types of cancer earlier than human radiologists (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). The application of AI in advanced screening, as well as in other detection technique, has demonstrated improved sensitivity and specificity compared to traditional methods. For instance, in a study conducted in the U.S., an AI algorithm screened with 91% accuracy cervical cancer, surpassing the 69% accuracy of human clinicians. A study on breast cancer detection through mammography screening, conducting at Capio Sankt Göran Hospital in Sweden and involving 55,581 women, concluded that double reading mammograms by one radiologist and AI algorithm has a cancer detection rate (0.5%) non-inferior compared to standard double reading by two radiologists (0.4%) (Dembrower et al., 2023; European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). This means that the combination of one radiologist and an AI algorithm is as effective as the standard practice involving two radiologists for cancer detection, but it occupies only one

tools may lead to inaccurate or misleading diagnoses and may discourage individuals from seeking appropriate professional medical advice, potentially posing risks to patient safety.

clinician, thereby improving the efficiency of procedures. In a study based on multiple datasets from China, U.S., and Germany, an AI model was found to outperform expert pathologists in diagnosing colorectal cancer (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). In Japan an AI system was developed to automate the process known as "Temporal Subtraction", which involves comparing medical images taken at different times in order to diagnose new bone metastases. This system enables radiologists to assess changes, which proves crucial given the complexity and urgency of identifying bone metastases.

In sum, most studies confirm the added value of using AI systems. These systems are often capable of identifying true positives with the same or greater effectiveness as human clinicians, thereby improving time and resource efficiency (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025).

3.1.3 Medical Robotics

Medical robotics are widely used since many years, as they allow very complex operations to be performed with greater precision. AI systems have been developed and deployed significantly to support robotic surgery, rehabilitation, and direct patient assistance.

AI is used to improve robotic systems that support surgical operations, through its implementation in robotic-assisted surgical systems and computer-assisted surgery. Robotic surgical systems—even in its basic form without AI systems—refers to robotic instruments, such as cameras and mechanical arms, that can be used to perform surgeries in confined spaces through small incisions with a high level of precision. Using these instruments facilitates operations and allows for a less invasive approach, resulting subsequently in faster patient recovery times and reduced postoperative complications (Väänänen et al., 2021). The use of robotic systems presents several benefits compared to open surgery, including reduced length of hospital stay, reduced blood loss and transfusion rates, and reduced complications (Ho et al., 2012). Instead, computer-assisted surgery refers to the computer technology used for surgical planning. It comprises several applications that assist surgeons at every stage of operations: preoperatively, intraoperatively, and postoperatively. Its features include the creation of virtual image of the patient, diagnostic, image analysis and processing, preoperative planning, surgical simulations, and surgical navigation (Väänänen et al., 2021).

The inclusion of AI algorithms in robotic and computer-assisted surgical systems can further enhance these tools. During the preparation stage of surgery, AI-supported analysis tools can analyze preoperative images for surgical planning purposes and anticipate possible complications based on the type of surgery and the patient's specific data. AI-driven robotic systems increase the precision of robotic instruments by supporting guidance during surgery and predicting complications, thereby reducing surgical errors. Their implementation also allows for the collection of a higher volume of data and details, such as video recordings of surgeries, and the movements and cutting and sewing actions during surgery, which can be used to enhance and facilitate the surgical process. Moreover, their implementation can mitigate the effects of factors such as fatigue and human inaccuracy, which can be critical in complex operations and confined spaces (Väänänen et al., 2021). An example is the robotic systems developed by Intuitive Surgical Inc and called Da Vinci, which is one of the most used in the field. Da Vinci is a surgical robot that allows clinicians to perform highly complex procedures with greater accuracy, flexibility and control respect traditional methods (Mudgal et al., 2022). The system translates the clinicians hand movements at the console and creates a 3D high-resolution image of the surgical site. In this way the experts dispose of an accurate image of the operation status (Khandalavala et al., 2020).

Alongside surgery, AI and robotics are also transforming approaches and competencies in rehabilitation research and practice. Research is underway into both physical (robotics) and virtual (computer science) applications of AI in rehabilitation. In this field, ML is used for perioperative medicine, brain–computer interface technology, myoelectric control, symbiotic neuroprosthetics, among others. ML methods have also been applied in the field of the musculoskeletal system. The Hybrid Assistive Limb (HAL) exoskeleton, developed by Japan's Tsukuba University and the robotics company Cyberdyne, is designed to help patients rehabilitate from lower limb disorders, like spinal cord injuries and strokes. HAL uses AI and bio-signals placed on the skin to detect electrical signals and respond with movement at the joint.

3.1.4 Virtual Assistance and Personalized Care

Among the various applications of AI in the healthcare sector, there is the development of virtual assistants. The creation of virtual assistants, in the form of chatbots or through analytical functions for symptom assessment, has been developing significantly, particularly after the Covid-

19 pandemic and with the expansion of telemedicine. Indeed, the pandemic highlighted the need for medical tools that enable healthcare professionals to work remotely, with patients remaining at home. Such tools have proven to be useful even after the crisis, allowing patients with limited mobility or those living in rural areas to conduct medical consultations and obtain test results without having to go to a clinic.

In the field of home healthcare, researchers have developed smart sensor system based on integrated sensor network to obtain data on a person's health status through the monitoring of a person's biomedical variables (Al Kuwaiti et al., 2023). These sensors often consist of active and sensitive wearable technologies that can measure physiological signals such as respiratory rate, pulse rate, breathing waveform, blood pressure, and ECG, and support patients in managing chronic conditions such as diabetes mellitus, hypertension, or sleep apnea (Kim et al., 2018). Since the outbreak of the Covid-19 pandemic, there has been significant progress in wearable devices that measure physiological changes in biometrics and transmit active patient monitoring online (Natarajan et al., 2020). All such systems permit HCPs to monitor and report patient conditions away from the traditional clinical settings.

Systems enhanced by AI have been developed to assist and support visually impaired individuals in their daily activities. RUDO, defined as an "ambient intelligent system," is an AI-based system designed for the domestic environment of a blind person. The system supports basic everyday needs and supplements the standard tools used by a blind individual in a household. It enables the recognition of incoming people and notifications about movement within the apartment, supports computer use—including specialized tasks in electrotechnics and informatics—and facilitates cooperation between a blind/sighted child with a blind/sighted parent (Hudec & Smutny, 2017). Designed for blind people, a similar system is also useful for other types of difficulties, such as dementia or even just elderly people who live alone at home.

Moreover, platforms that take the form of virtual assistants and can interact with individuals—asking about their symptoms and needs and providing an initial consultation—are becoming increasingly widespread. Through ML and NLP, these virtual nurse assistants can listen, converse, and provide recommendations to patients, helping clinicians better understand their cases. Studies on the use of embodied conversational agents in healthcare have demonstrated significant improvements in health outcomes when chatbots or conversational agents are employed. These systems contribute to reducing the workload of healthcare professionals and provide

assistance both in hospital settings and within home environments. The main providers are GYANT, Babylon Health, Ada Health, and Sensely. All of them offer AI systems to address symptom triage, mental health assistance, and chronic disease management.

One widely used platforms is Your.MD, which employs AI and ML to deliver personalized pre-primary care. The platform utilizes U.K. National Health Service data to assist patients before they decide to access primary care services. Through a mobile app or website, users can interact with a chatbot, describe their symptoms, receive an initial assessment, and be guided to contact specialized clinics or hospitals. Benchmark tests in verified cases from Harvard University and Royal College of General Practitioners have shown medical accuracy of 85% for 20 most common conditions. Additionally, the system provides safe urgency advice with 92.6% accuracy (Väänänen et al., 2021). Similarly, GYANT chatbot assists patients in understanding their symptoms. The data and symptomatology collected by the system are then transmitted to physicians, who can diagnose conditions and prescribe medications in real time. In 2024, GYANT chatbots are being used by over fifty hospitals around Europe for preliminary patient screenings. Sensely also has launched virtual assistant chatbots. The platform, based on ML technology, consists of a digital nurse avatar that can interact with patients and facilitate condition monitoring. The platform can interact with medical devices, such as blood pressure monitors, while also taking patients' medical history data into account. In this way, the platform supports nurses in their work by keeping healthcare providers and patients continuously connected, thereby reducing the workload and pressure on medical staff. Moreover, Sensely can also keep track of appointments, bridge the gap between doctor visits, and recommend follow-up treatments. The platform was trialed in 2019 with 72 patients with chronic heart failure at a clinical site. The findings showed that the platform decreased readmission rates by 75% and monitoring costs by 66% versus the traditional care process (Väänänen et al., 2021; Mudgal et al., 2022). AI apps that monitor and assist patients in following their prescribed medication and treatment plans have also been shown to be effective.

Chatbots focused on mental health have been developed. Woebot is a mobile app that uses Cognitive Behavioral Therapy to listen and offer advice to anyone who seeks it out. It asks people how they feel through short daily conversations and sends advice depending on the mood of the moment and how the person responds to the questions. The creators' goal was to encourage people––under 30 in particular—to engage in therapeutic conversations by breaking down the social stigma surrounding distress. A study conducted with Stanford University showed that 85% of participants

aged 18 to 28 who used Woebot daily reported a significant reduction in anxiety and depression symptoms (Fitzpatrick et al., 2017).

The rise of virtual assistance significantly reshapes healthcare delivery procedures by expanding territorial coverage and improving access to care. In theory, these systems enable patients to receive medical guidance, monitoring, and preliminary assessments remotely, thereby reducing the need for physical visits to healthcare facilities. This capacity becomes particularly crucial during health emergencies, as demonstrated during the Covid-19 pandemic, when remote consultations and digital triage systems helped maintain continuity of care while limiting physical contact. The possibility of managing high volumes of requests from home not only enhances system efficiency but also contributes to reducing the spread of infectious diseases. Nevertheless, the effectiveness of such systems ultimately depends on their quality and reliability. Poorly designed or inadequately supervised virtual assistance tools risk generating misinformation, diagnostic errors, or unequal access, potentially creating new vulnerabilities rather than alleviating existing pressures. Ensuring the robustness and accountability of these systems is therefore essential.

3.1.5 Drug Discovery

Drug research and discovery is another area with significant potential for AI applications. The processes of new drug discovery are typically long, complex, and costly, with numerous potential points of failure. AI's analytical capabilities can support these processes, helping to reduce both the high costs and the lengthy timeframes required for new drug development. The emergence of epidemics and pandemics, such as influenza and Covid-19, as well as the ongoing prevalence of severe diseases such as cancer and cardiovascular disorders, has highlighted the continuous need for the discovery of new drugs.

AI can prove useful at various stages of the process of developing and implementing a drug in clinical settings. Beginning with the target identification phase, in which molecules capable of counteracting a specific disease state are identified, AI can support the analysis of data such as molecular interactions in order to identify potential targets that are likely to be involved in disease pathways. In virtual screening and optimization of compounds procedures, AI can be used to virtually screen and optimize compounds to estimate their bio-activities and predict protein-drug interactions. The use of AI predictive models would make it possible to identify compounds with

a high probability of binding to a target. AI can be deployed to plan efficient chemical synthesis routes and gain insight into drug reaction mechanisms to identify potentially unwanted interactions with other molecules. NuMedii, a Biopharma firm, developed an AI system for drug discovery able of identifying connections between drugs, diseases, and systems (Mudgal et al., 2022). In the pre-clinical stage, which involves predicting possible responses to a drug, AI—specifically ML and DL—can provide support through similarity and feature-based methods. Similarity-based methods assume that similar drugs act on similar targets, while feature-based methods identify individual features of drugs and targets and feed the drug-target feature vector to the classifier. Finally, AI systems can also help in the selection of potential patients for pre-clinical trials by identifying pertinent human disease biomarkers and anticipating potential toxic or harmful side effects (Qureshi et al., 2023).

AI systems in drug discovery can therefore prove extremely valuable by enabling the analysis of vast amounts of biomedical data and accelerating research processes. This capacity support robust research and development, resulting in significant financial benefits and enhanced health security.

3.2 Opportunities and Challenges of AI Integration in Healthcare

AI, ML and DL applications in healthcare have increased significantly in recent years. Numerous research institutes, private companies and universities worldwide are investigating the use of these technologies in the healthcare sector. The number of EU-funded research projects has grown substantially, particularly between 2019 and 2022, driven in part by the Covid-19 pandemic, which highlighted the essential need for a strong, resilient, and technologically advanced healthcare system (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). By mid-2025 the list of the U.S. Food and Drug Administration (FDA) of the approved AI/ML-enabled medical devices has counted over 1250 devices. All devices included in this list have met premarket requirements, including an assessment of their safety and effectiveness (Food and Drug Administration (FDA) (U.S.), n.d.). Nevertheless, despite the increase in research projects and AI-based medical devices, the effective deployment of AI systems in clinical practice remains limited. A substantial discrepancy persists between the volume of research and development devoted to AI-enabled medical devices and their actual adoption in routine clinical settings.

As in the military domain, the implementation of such technologies in the healthcare sector is accompanied by an extensive debate on the technical challenges, as well as the ethical and social implications, associated with their use.

3.2.1 Technical Challenges

A first significant challenge concerns the datasets needed to train the algorithm for the healthcare field. AI systems to function effectively, they require large volumes of data on which to be trained. In this regard, AI technologies prove to be particularly effective in fields where abundant data are available for training—as has been seen in radiology (see section 3.1.2)—while perform poorly in contexts characterized by data scarcity or difficulties in digital data collection. If poorly designed, models are more likely to develop biases or other types of errors and, consequently, produce harmful or discriminatory outcomes. Moreover, instead of reducing the workload of HCPs, they may increase it due to the need for continuous monitoring, verification, and correction (OECD, 2024*a*). For instance, in a Japanese hospital representative pointed out that using AI in diagnostic imaging could increase the workload of radiologists by requiring them to review a greater number of false positive results (OECD, 2024*a*; European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). One of the causes of the scarcity of large datasets in healthcare is data heterogeneity, as it limits its integration and interoperability. The lack of standardized data structures negatively affects AI's ability to effectively analyze and aggregate information across different systems and requires complex data transformation processes to ensure interoperability. For instance, to integrate an AI model in oncology, it was necessary to develop a tool capable of converting the output generated by the AI system into a format readable by HER (Gao et al., 2024). This limits the resources available to train, refine, and test AI algorithms, ultimately constraining their overall performance, as well as the effective implementation of AI in healthcare.

Another issue concerns the "black box" nature of AI and the difficulty in interpreting and tracing the decision-making pathways that lead to the outcomes proposed by the algorithm. Clinical decisions must be grounded in clearly interpretable scientific evidence enabling practitioners to assess their validity and, importantly, to justify clinical choices and assume responsibility for their outcomes. The lack of transparency and explainability seems to contradict the principles of

evidence-based medicine, as healthcare professionals may struggle to evaluate, validate, and appropriately apply AI-generated recommendations in clinical practice. As already discussed in this thesis, being designed by humans, AI systems inherently carry the possibility of errors and biases reflecting those of their designers. As a result, even in the healthcare domain, the risk of erroneous outcomes remains present. For instance, an AI application used to predict the likelihood of patients developing complications after pneumonia incorrectly advised physicians to discharge asthmatic patients (Al Kuwaiti et al., 2023; Mudgal et al., 2022). The inability to explain AI systems is an important issue that must be addressed to enable their safe implementation also in healthcare. Moreover, this issue becomes particularly critical when considering that AI models are typically developed by non-medical professionals; as a result, end users—healthcare providers and patients—have limited control over how results are generated. This not only poses challenges during the development phase, which requires close collaboration between domain experts and developers, but also raises significant challenges for government policymakers and regulators.

Finally, there is a technical difficulty associated with the infrastructure needed for the deployment of such systems in hospitals. Indeed, the implementation of AI systems requires adequate IT infrastructure capable of supporting them, as well as a continuous training for hospital staff to ensure they know how to use it fully and effectively. Unfortunately, various hospital facilities operate with obsolete IT infrastructures due to limited financial resources. Especially in Europe, many healthcare facilities do not have digital EHRs. Often, they operate with systems not able of supporting the advanced computational requirements of AI technologies, as they lack the necessary processing power, storage capabilities, and network bandwidth needed for AI applications (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). Indeed, AI technologies entail high costs related to the acquisition and maintenance of the necessary computing infrastructure, but also additional expenses for data storage and backup. The efficiency of these processes must be high in order to ensure continuous system updates and adequate data protection. When combined with the economic and time-related costs required for the training of healthcare professionals, the implementation of such technologies represents a particularly significant financial burden for hospitals. The cost–benefit balance of expenditures related to the implementation of such technologies tends to be achieved in the long term through faster clinical consultations, shorter hospital stays resulting from less invasive surgical procedures, and the more efficient optimization of human and physical resources.

3.2.2 Regulatory and Legal Challenges

At the legal level, the high degree of regulation in the healthcare sector, combined with the uncertainty surrounding regulatory interpretation, makes the evaluation and approval of AI applications particularly critical. In fact, the implementation of AI technologies in healthcare must be accompanied by adequate governance frameworks in order to address regulatory, ethical, and trust-related challenges.

Concerns about data privacy and security are highly relevant. The collection and processing of large amounts of sensitive patient data in large-scale datasets raises several accountability and regulatory concerns. Questions arise about where and how the data processed by AI solutions is stored. Many AI solutions rely on cloud-based platforms and may require data to be transferred and stored across different jurisdictions, potentially outside the EU. This raises urgent concerns about compliance with the legal framework on data protection, especially in regions with weaker data protection standards and the risk of unauthorized access (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). In addition, there is a risk arising from the lack of full transparency regarding the use and the potential misuse of the data collected by AI tools. Beyond diagnostic or strictly medical purposes, these data could be used for secondary purposes, such as commercial profiling or research activities, without an appropriate legal basis. This risk is exacerbated by limited transparency in how some AI solutions manage data after deployment, thereby creating challenges in maintaining patient trust. Finally, there is the risk of cyberattacks and data breaches. Given the highly sensitive nature of the data involved, the risk of cyberattacks and malicious activities is significantly heightened, thereby requiring particularly robust and carefully designed security measures. In fact, since the Covid-19 pandemic, cyberattacks targeting hospitals and healthcare systems have significantly increased, resulting in significant financial losses and serious health-related consequences, including breaches of patients' medical records and disruptions to critical services (Wasserman & Wasserman, 2022; Mudgal et al., 2022; European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025).

Added to these points is the central issue of liability. In the same way as in the military, questions arise concerning responsibility for decisions made by AI systems. It becomes evident that there is a need to determine who should be held accountable in the event of AI-related errors–

–whether the developers, the clinicians, or others—and, consequently, who would be responsible for providing compensation to individuals adversely affected by the use of such technologies. This connects to the challenges of transparency and explainability and raises important legal questions for the implementation of such tools.

3.2.3 Socioeconomic Considerations

On a socioeconomical level, AI systems have the potential to improve the allocation of healthcare resources avoiding waste and supporting a more efficient and equitable distribution. Especially in isolated, peripheral areas, using AI models to make more strategic and efficient use of resources can help maximize the utility of available assets, enabling patients to avoid excessive delays or shortages in critical healthcare services (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). Such systems can help optimize the distribution of HCPs by analyzing available personnel, patient needs, and transportation logistics to determine the optimal placement of mobile clinics and schedule rotating specialists to serve multiple remote communities (Kong & Shaoshan, 2022). Their ability to improve diagnostic services and overall hospital times, combined with better visit administration, can contribute to healthcare delivery. Predictive AI models enable for better preparation for seasonal fluctuations in healthcare demand (Dixon et al., 2024).

AI technologies, by enhancing healthcare delivery, diagnostics, and operational efficiency, can contribute to mitigate the issue of widening disparities in access to healthcare and bridging healthcare access gaps, particularly for populations living in rural and underserved areas. Indeed, in many rural areas, the scarcity of healthcare resources—often geographically distant—constitutes a major barrier to access (European Commission: Directorate-General for Health and Food Safety, EIGE, Open Evidence & PwC, 2025). Additionally, the rise of virtual assistant models, such as those discussed above (see section 3.1.4), can reduce the need for frequent in-person consultations, which alleviates the demand on clinics, improves resource management and reduces transportations costs for patients. Telemedicine—especially since the onset of the Covid-19 pandemic—has emerged as an important tool for making healthcare more accessible and AI technologies can further improve these services (Sharma et al., 2023).

Moreover, with the aim of improving healthcare service provision in geographically disconnected areas, AI can play an important role in training and upskilling local healthcare providers. AI can simulate clinical scenarios, teach new diagnostic methods, and offer insights based on real-world data through virtual training modules. AI-enabled training platforms use realistic simulations to help healthcare providers practice procedures, learn about new treatments, or refine diagnostic skills. In rural areas, where physicians may not have the same possibilities of access to specialty, AI-assisted training can help fill knowledge gaps (Hamilton, 2024).

3.2.4 Ethical Challenges

On an ethical level, one of the main obstacles to the implementation of AI is the lack of transparency. The inability to fully understand how these systems operate undermines trust in them and, consequently, their adoption. AI systems are often introduced without adequate explanations of how they work or how decisions are generated, which leaves users confused and skeptical. The issue is closely linked to the "black-box" nature of AI systems and, consequently, to the lack of transparency—in a sense justified. In this case, trust should be built through clear and effective communication among developers, deployers and end-users, greater transparency on the part of relevant stakeholders, and the use of extensive demonstrative testing to validate and illustrate system performance (OECD, 2024*a*; European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025).

Another issue is the level of knowledge and literacy in the medical community regarding digital technology and AI. Many HCPs don't have the basic knowledge or skills to effectively engage with these digital tools, including understanding how these systems work, their potential applications, and their limitations. As a result, clinicians may feel unprepared to use the technology in an appropriate manner or may distrust its outputs. It is unlikely that healthcare professionals will incorporate these tools into their clinical workflows if they do not have a fundamental understanding of how to interpret and use the results of the algorithms. Using AI without adequate training could pose potential risks also to patient safety as a result of an improper way of interacting with the system, such as erroneous input, or misinterpretation of the output. Therefore, continuous training and refresher courses are needed for HCPs so they can use AI tools with a mastery of the

system's potential and an awareness of its limitations (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025).

The widespread diffusion of medical AI models could enable individuals to engage in self-diagnosis and manage certain healthcare needs autonomously at home, without consulting a physician. While this could help increase individuals' health awareness and provide a rapid, easily accessible initial medical consultation from home, it also carries the risk of negatively impacting patient autonomy and leading to isolation (Mudgal et al., 2022). Without a broader clinical assessment that includes a human physician, patients risk narrowing their treatment options and consequently limiting their ability to provide fully informed consent regarding medical procedures. For this reason, even though these systems may offer valuable consultations, subsequent evaluation by a human healthcare professional is necessary at the current stage of AI development.

Finally, an important point of discussion is the consequences of the implementation of such systems on the doctor-patient relationship. The relationship between doctor and patient is highly important in a sensible sector such as healthcare. This relationship is usually characterized by mutual trust, effective communication, empathy, and collaboration, dealing with health and possibly life of the patient. HCPs trust the information provided by patients, while patients, in turn, place their trust in the expertise of their healthcare providers, relying on their guidance for accurate diagnoses and effective treatments. Particularly in severe and delicate cases, it is the human relationship that supports patients in coping with a diagnosis or undergoing treatment. Although healthcare professionals are sometimes portrayed as being dehumanized by their work, in practice the human dimension remains present and is essential.

The large-scale integration of AI in the healthcare sector could create a sense of alienation between patients and healthcare professionals. The essential human dimension of patient care may become less prominent as AI systems assume greater responsibilities in clinical decision-making and care delivery, potentially undermining patient trust. Patients presenting AI-informed information—whether accurate or misleading—may complicate the process of shared and collaborative decision-making with physicians. Indeed, aside from the issue of insufficient trust, there is also the risk of excessive trust. Patients may place greater confidence in assessments generated by AI systems even when medical specialists question them, thereby creating potential conflicts and further undermining trust in human experts.

The implementation of AI in healthcare is therefore neither simple nor linear. When well designed, such systems can represent significant added value; however, if poorly designed, they entail numerous risks and may complicate physicians' work. For instance, a HCP from the U.K. expressed concern that some diagnostic AI tools might hinder experienced HCPs, leading them to second-guess themselves (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). For these reasons, providing adequate information about these tools is crucial. As highlighted several times in this thesis, AI systems are intended to support human activities rather than replace it. In order to avoid risks, gradual integration accompanied by appropriate information for HCPs and patients would be necessary. It is essential to not forget that human expertise remains at the center of healthcare processes.

3.3 The Concept of Health Security and AI-enabled Healthcare

What are the reasons for treating health as a security matter, and what would be the implications of a strong—AI-enhanced—healthcare sector? To frame this discussion, it is necessary to introduce the concept of *health security*, a term that remains fluid and contested in both academic and policy debates. Indeed, while the link between the military sector and security appears evident and direct, the connection between healthcare and security is far less visible and tangible, although not for this reason less significant.

The recent Covid-19 pandemic highlighted such connection, showing the relevance of health to national and global security and triggering unprecedented restrictions on civil liberties. In several countries, including France, Spain, and Italy, military forces have been activated in response to the crisis and operated as key responding actors by setting up field hospitals and enforcing lockdown measures. Part of the academic community has drawn increased attention to the elevated role of law enforcement agencies during the Covid-19 crisis and to the extraordinary expansion of policing powers—examples of this securitization of the health emergency. However, the securitization of health predates the Covid-19 pandemic. The rising costs associated with chronic diseases, obesity, and the current opioid crisis in the U.S. have been identified as national security challenges, contributing to the broader spectrum of health-related security issues (McCoy et al., 2023). Migration flows have been framed within the context of health security, with unregulated flows of people being portrayed as a threat to local health systems. The unpredictable health emergencies resulting from terrorist attacks have highlighted the need of health systems able

of dealing with such scenarios, framing the issue as a matter of health security. The 2014–2015 West African Ebola epidemic marked a critical moment in framing disease as a foreign policy issue and an international security concern. Going further back, toward the end of the 20th century, the spread of HIV/AIDS in several African countries was identified by the U.S. Department of State as a security threat (McCoy et al., 2023). This paragraph, therefore, aims to contextualize the concept of health security in order to better frame the connection between the two, as well as the implications of maintaining a strong healthcare system.

As already mentioned, there is no consensus on the application of the concept of security to the health domain, with ongoing debates over the usefulness of framing health as a security threat in the same way as terrorism or nuclear proliferation. Critics highlight that such a security-oriented discourse may not be well suited to addressing public health challenges. Nevertheless, developments over the past decades appear to confirm the existence of a connection between health and security. The World Health Organization's (WHO) International Health Regulations (2005) reflect a securitization of health by encouraging states to strengthen their capacity to "detect, assess, report, and respond to potential public health emergencies of international concern". The WHO defines health security as "the activities required, both proactive and reactive, to minimize the danger and impact of acute public health events that endanger people's health across geographical regions and international boundaries" (WHO, n.d.).

Over time, two alternative and mutually tensioned conceptualizations of health security have emerged. The first emphasized the security of healthier and wealthier countries—typically former colonial powers—and frames poorer countries as a threat because potential sources of disease and epidemics. This approach has been described by McCoy et al. (2023) as *neocolonial health security*. On the other hand, a second approach adopts a more inclusive prospective, incorporating the needs of all populations and viewing the health conditions of poorer and disadvantaged groups—such as poverty, hunger, limited access to healthcare, and human rights violations—as central issues to be addressed. This perspective is referred to by McCoy et al. (2023) as *universal health security*, underling the globally oriented vision. The key distinction between these alternative conceptualizations of health security lies in which area of global society that is prioritized. The first approach focuses on protecting the security of wealthier populations and aims to isolate and contain the consequences of poverty and diseases in poorer countries. Universal health security, on the other hand, recognizes the needs of low-income populations as a root cause

that must be addressed to achieve full health security. The dominant discourse tends to oscillate between these two conceptualizations, dividing scholars and policymakers between those who advocate for the need for an open and genuinely global approach to health security and those who emphasize the associated challenges, arguing that even a wealthier-countries–privileged approach to health security can lead to trickle-down effects. The neocolonial health security approach frequently prevailed. During the West African Ebola epidemic, vast amounts of funding and effort were directed toward preventing the entry of the disease into Northern and Western countries under the declaration of a Public Health Emergency of International Concern (PHEIC), while affected populations in West Africa faced severely under-resourced health systems. However, this approach is somewhat limiting in today's world. In an increasingly globalized and interconnected world, it is no longer possible to overlook segments of global society. Threats—including health-related threats—have become increasingly transnational and highly networked in nature, making inclusive approaches that account for all affected actors necessary and unavoidable.

As Akhavein et al. (2025) underline, health has never been *un*-strategic. From a socioeconomic perspective, a deadly epidemic has the potential to disrupt global supply and value chains, revenue streams, and international trade. The high level of interconnectedness and international exchange significantly increases the risk that an infection may spread rapidly and affect the global system as a whole, thereby posing a social and economic threat not only to the country directly affected but also to those states that are relatively untouched by the disease itself (McCoy et al., 2023). HIV/AIDS was identified as a security threat in 2000 due to its potential to destabilize societies and economies. The same reasoning has been applied to outbreaks such as SARS and, more recently, Covid-19. Bioterrorism and fabricated health emergencies represent serious security threats (Wright, 2006; Akhavein et al., 2025). Biological weapons were experimented with and used during the First and Second World Wars, and more recently in events such as the 2001 anthrax mailings in the U.S., in which, less than a month after the September 11, 2001, attacks, letters laced with anthrax were sent to media offices and to two U.S. senators, resulting in five deaths and the infection of twenty-two individuals. Such events have reinforced the perception of health protection as a vital national security measure.

Healthcare systems exert a profound influence on broader dimensions of social security. This is exemplified by the long-standing debate between public and private healthcare models. Private healthcare systems, by conceptualizing healthcare as a business activity, tend to priorities

efficiency, technological advancement, and service quality. However, such systems carry significant social consequences, as access to care is largely restricted to economically privileged groups, which contribute to increased social discontent, the amplification of existing inequalities, and effects on phenomena such as social dependency and deviance. By contrast, public healthcare systems tend to emphasize human security by prioritizing equity and universal access, thereby fostering social stability and collective well-being. The consequences of these differences are relevant. For instance, the U.S. government's ability to detect and respond to disease outbreak is constrained by structural inequalities within the healthcare system, such as gaps in health insurance coverage for a significant segment of the population and restricted access to federally funded healthcare programs for undocumented migrants (Gutlove & Thompson, 2003). Of course, economic considerations are also critical in such discourse. While private healthcare systems tend to be more financially sustainable and flexible, public systems, which rely heavily on state funding and are vulnerable to economic downturns, often face structural inefficiencies and financial strain. This is intended to shed light on how the structure of a healthcare system directly affects human security and how structural inequalities in access to care can translate into broader vulnerabilities that undermine societal resilience and national stability.

Finally, from a strategic and geopolitical perspective, the securitization of health-related issues may pose a threat to a state's internal stability. The securitization of a health event or crisis entails the attribution of a sense of urgency and exceptionalism, which can lead to the adoption of extraordinary measures—such as travel restrictions, military-enforced lockdowns, and mass surveillance—that fall outside the realm of "normal politics," as was seen during the Covid-19 pandemic (Akhavein et al., 2025). In such cases, the pressures at play and the balance between the public health needs and the geopolitical interests can appear fragile.

It follows that a well-functioning healthcare system is crucial to country's overall well-being and security resilience. In this context, AI can play a positive role by enhancing the efficiency and effectiveness of healthcare systems, improving both their capacity to respond to health crises and their ability to provide healthcare services that are more accessible and efficient.

Through data-driven optimization, AI enables more efficient allocation of physical and financial resources and thereby alleviating strain on healthcare facilities and personnel (European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC, 2025). AI can support healthcare providers in diagnostics and treatment optimization, in this way

accelerating and improving clinical procedures and responding to high levels of demand (OECD, 2024*a*). These technologies provide valuable assistance to HCPs, allowing them to devote more time to patient engagement by reducing the administrative burden—an estimated 36% of activities in the health and social care sector are considered potentially automatable through AI (OECD, 2024*a*)—thereby partially addressing staff shortages and helping prevent the collapse of care delivery systems. AI also accelerates the drug discovery process, making it faster, more cost-effective, and more efficient, thereby contributing to greater drug availability, enhanced population-level resilience to disease, and financial benefits derived from a stronger research and development sector. As the OECD (2024*a*) notes, *"*AI can help the health sector to unlock value from the 97% of health data assets that are not currently used to assist decision-making*"*. Collectively, these developments can translate into an overall improvement of healthcare systems, enhancing more preventive healthcare, with positive spillover effects that extend beyond healthcare.

Moreover, cyberattacks are increasingly targeting hospitals and healthcare systems—according to the European Commission, "the healthcare sector is one of the most targeted by cyberattacks" (European Commission, n.d.*c*)—resulting in serious financial losses and disrupting the provision of healthcare services. These attacks undermine national resilience by weakening emergency response capacities and represent a growing public health and national security concern (Wasserman & Wasserman, 2022). In this regard, AI can also play a defensive role by supporting the detection of cyber threats and contributing to the prevention. When appropriately governed and integrated, applications may strengthen trust in the secure use of digital health technologies (OECD, 2024*a*).

Beyond routine healthcare activities, AI can significantly enhance a state's capacity to manage health crises by improving preparedness and responsiveness. AI-based early warning systems can identify emerging health risks, while AI's ability to rapidly process, analyze, and synthesize information under time pressure improves crisis management capabilities (Corrigan, 2024; WHO, n.d.*b*). For instance, the Epidemic Intelligence from Open Sources (EIOS) enables the rapid detection of public health threats and leverages NLP and ML techniques to identify public health events from open-source information using an all-hazards approach (Williams et al., 2025). Tools such as the WHO All-Hazard Information Management Toolkit, designed to support emergency information management, leverage AI to enhance crisis response through the production of rapid risk assessments, response plans, monitoring tools, and situation reports (WHO,

n.d.*b*). Additionally, AI can support disease surveillance, medical workforce management, and drug discovery, while AI-enabled healthcare logistics facilitate optimal resource allocation, minimizing the duration and severity of disruptions. Taken together, these applications enable states to detect potential risks in advance and to respond in a more effective and coordinated manner during health emergencies, fostering a more proactive and resilient healthcare security.

4. Cross-sectoral Challenges in the Development and Governance of AI

The development of AI in both the defense and healthcare sectors has highlighted the remarkable adaptability of this technology and its applicability across seemingly distant domains, stemming from its nature as a dual-use technology that enables such broad potential. Although these sectors differ significantly in purpose and operational context, AI emerges in both as a powerful tool, improving the efficiency and effectiveness of several processes. By analyzing vast quantities of data within relatively short timeframes, AI systems enable the extraction of greater value and actionable insights from available information. This capability offers notable advantages by enabling higher levels of informational availability and greater analytical depth. In the military domain, enhanced data-processing capabilities support intelligence analysis and decision-making, while in health secuirty, it allows for the extraction of significantly greater value and usability from health data assets. The capacity of AI systems to monitor data streams and identify patterns and anomalies significantly expands analytical possibilities, highlighting trends that could have remained undetected by human operators and allowing early warning mechanisms. In the defense domain, it supports the monitoring of network of flows to detect cyberattacks and the identification of suspicious behaviors in online counterterrorism activities, thereby contributing to more effectively preventing these novel hybrid threats posed by organized crime and hostile states. In healthcare settings, it allows for improved prevention and timely treatment—a capability that emerges as particularly important in light of the current demographic trends, which place increasing pressure on healthcare systems and heighten the need for proactive and preventive approaches. Additionally, AI predictive and anticipatory capabilities allows to estimate outcomes, optimize operations, and anticipate potential challenges on the basis of available data. This represents a unique asset, which, in the military, is reflected in applications such as the analysis of possible course of action and the predictive maintenance of equipment, while in healthcare settings it supports the management of patient flows and emergency departments.

Collectively, these applications contribute to a state's responsiveness, anticipatory and resilience-oriented abilities. Through such improved analytical capabilities, AI supports better preparation and faster response, therefore supporting a shift toward more proactive approaches (Misuraca & Noordt, 2020). These capabilities translate into a more prepared security position in the defense domain, characterized by greater interoperability, improved analytical capacity, and

faster response mechanisms. While, in the healthcare sector, it enables systems that are better equipped to develop responses more rapidly and with greater resilience, improving health outcomes and people-centered care and enhancing health security (OECD, 2024*a*).

However, these developments also signal a deeper transformation. The integration of AI does not only improve existing processes but has the potential of reshaping them, modifying decision-making procedures and redefining human role. Indeed, AI integration into military structures seems to suggest a structural transformation of procedures and responsibilities. The data-processing capabilities of AI accelerate the OODA loop, enhance predictive targeting, and automate elements of intelligence analysis and decision-making. The rise of hybrid approaches—such as AI-enabled cyberattacks, information manipulation, and influence operations—modifies the methods and intensity of destabilization strategies. In doing so, AI technologies reshape human role and nature of control. As autonomous systems assume greater roles, human involvement becomes more supervisory and less directly engaged. Rather than executing and evaluating actions, human operators increasingly supervise systems that generate outputs, analyses, or recommendations, with control often exercised through centralized command structures. Decision timeframes are compressed and the space for human deliberation is narrowed. This shift raises questions considering the risks of misinterpretation, algorithm bias, technical malfunction, or cyber manipulation—particularly when human oversight becomes procedural rather than substantively engaged. Appropriate security mechanisms to ensure control, accountability, and human-machine relationship must be considered. Moreover, growing reliance on interconnected infrastructures increases vulnerability to disruption and interference. Consequently, while such developments generate considerable tactical advantages, they also require a rethinking of operational roles and security frameworks in order to address these emerging concerns and ensure AI enhances security rather than creating vulnerabilities.

Similarly, in healthcare systems, the integration of AI reflects a dual dynamic. One the one hand, AI enhances states' capacity to anticipate, prevent, and respond to health-related shocks in a more effective and coordinated manner, thereby strengthening healthcare resilience and preparedness and contributing to enhanced human security. Through diagnostic support, improved analytics, early-warning systems, and optimized resource allocation, AI promotes a more responsive and adaptive health security approach—which appears particularly crucial in light of the demographic trends illustrated. On the other hand, the growing reliance on AI-based

infrastructures introduces new vulnerabilities. System errors and algorithmic biases may compromise medical services, while increasing digitization exposes healthcare infrastructures to cyber threats and privacy issues. The rise in cyberattacks targeting hospitals illustrates how healthcare has become an increasingly strategic and critical domain for human security and national stability. Consequently, also the securitization of healthcare infrastructures calls for approaches capable of addressing the emergence of new threats.

In this evolving landscape, failure to adopt these technologies may translate into a disadvantage and increased vulnerability to emerging threats. However, their adoption also entails a number of challenges. The development and integration of AI-based security capabilities creates dependencies on external resources—including advanced semiconductors, data infrastructures, and private technology providers—that may expose states to supply chain limitations or strategic leverage by third actors. Moreover, the increasing reliance on AI systems raises risks—such as the loss of control deriving from AI's opacity—that underscore the necessity of robust safeguard mechanisms. Addressing these challenges requires a comprehensive strategy that combines technological investment with institutional regulation and diplomatic negotiation. This includes the development of secure supply chains through inter-state cooperation, as well as an ethically grounded approach to governance that considers responsibility, accountability, and long-term implications.

The following section will examine the cross-sectoral challenges that have emerged in greater depth, with the aim of analyzing the broader implications associated with the widespread implementation of AI and outlining pathways for its responsible and secure deployment.

4.1 Transparency, Accountability and Control Issues

Both sectors have highlighted some significant concerns related to the opacity of AI systems, the risk of loss of control, and the possibility of responsibility gaps, which pose considerable security risks when such systems are introduced into highly sensitive domains.

The lack of transparency and AI's "black box" nature emerged as a key issue in both case-studies, particularly given the criticality of entrusting responsibility to a system without having a clear understanding of how it evaluates information, concludes decisions, and act in dynamic environments. Consequently, concerns arise regarding reliability, trust, and controllability. In the

defense sector, reliance on opaque systems may contribute to miscalculation, unintended escalation, or erroneous targeting decisions. Similarly, in healthcare, a field grounded in exact sciences, reliance on non-transparent algorithmic recommendations may increase the risk of misdiagnosis and patient harm. In both cases, opacity complicates effective oversight and hinders the timely detection of malfunctions, bias, or adversarial manipulation.

In response to these challenges, major institutions increasingly advocate a risk-based approach to AI governance, which differentiates regulatory requirements according to the potential severity of harm. This approach is exemplified by the European Union AI Act (Regulation (EU) 2024/1689). Risk assessment and differentiated governance can also be found in the policy frameworks developed by NATO and the OECD (NATO, 2024; OECD, n.d.*b*). A risk-based approach appears as necessary and such frameworks represent an important attempt to reconcile innovation with precaution. Nevertheless, the limitation of the existing provisions lies in the fact they differ across jurisdictions and are mostly politically negotiated. The absence of internationally accepted standard for defining and assessing AI risks may hinder the effectiveness of such provisions, particularly considering the transboundary nature of AI systems industry and of AI-generated harms. The development of common risk-based guidelines could be important for an effectively safe implementation of AI systems. In high-risk domains tolerance for uncertainty must remain minimal and deployment should be accompanied by continuous and rigorous testing, validation, and training processes, as well as robust human oversight, in order to ensure reliability, accountability, and safety.

Both sectors have also emphasized the issue of accountability, specifically the need for mechanisms for determining and ensuring responsibility for AI-driven outcomes and decisions. If such systems are to assume an increasingly prominent role—directly or indirectly—in decision-making processes, mechanisms must be in place to ensure accountability, as responsibility cannot be attributed to the machine itself. Complex chains of agency emerge, involving developers, deployers, operators, institutions, and regulators. This diffusion can generate dangerous responsibility gaps. To prevent accountability gaps, emerges as essential to establish frameworks that clearly allocate roles and responsibilities throughout the AI lifecycle and to clarify whether responsibility lies with the organization, the individual, or both in the event of errors or harm. To this end, NATO has underlined the need of meaningful human oversight, as has the European Commission in its *Ethics guidelines for trustworthy AI*, subsequent policy documents, and the EU

AI Act (NATO, 2024; European Commission, 2019; Regulation (EU) 2024/1689, art. 14). Meaningful human presence throughout the entire AI lifecycle therefore emerges as essential to assess, assign, and enforce legal responsibility. Human oversight, however, cannot be assumed to function automatically as a safeguard, as risks of automation bias, deskilling, and over-reliance on algorithmic outputs—particularly in high-pressure operational environments—are present. Therefore, human presence must be substantive and not merely formal. One possible precautionary measure could involve periodic testing and evaluation of the human operators responsible for supervising such systems, in order to ensure that their control remains effective and informed. Moreover, the form of human–AI interaction varies across sectors, and also this variation must be taken into account in governance discussion. In healthcare, human–AI interaction is likely to take the form of collaboration, with humans and AI systems working together, for instance, in diagnostic processes. A cross-sectoral collaboration between technology developers and HCPs is needed to ensure that clinicians are adequately prepared and trained to engage with these systems, use them effectively, and critically supervise their outputs, thereby preventing harmful consequences and ensuring patient safety. In the military domain, by contrast, the cases analyzed suggest that human involvement is more likely to take the form of supervision and oversight, where time constraints and operational stress may further challenge meaningful intervention. Despite the different modalities of human–AI interaction, in both contexts human agency must remain informed, competent, and grounded in an adequate understanding of the system's capabilities and limitations (Csernatoni et al., 2025), in order to avoid situations in which humans follow machine guidance without thoroughly supervising outcomes.

Therefore, transparency and accountability represent challenges for maintaining control over AI systems and ensuring their safe implementation. These are interdependent dimensions, as if opacity limits understanding, appropriate accountability mechanisms must be established to guarantee a strict control throughout the AI lifecycle. Such mechanisms should be grounded in human oversight, which must be carefully studied to ensure it remains substantive and meaningful. Although strategic competition is pushing for accelerated deployment, safeguards must be put in place to prevent the erosion of control over the technology and to avoid AI becoming a new source of systemic vulnerability rather than enhanced security.

4.2 Data, Resources, and Strategic Dependence

The analysis also highlights structural vulnerabilities in AI development that have direct implications for national and international security. As AI increasingly becomes an essential component of states' security architectures—both directly and indirectly—the ability to access and mobilize key enabling resources, including data, computing power, and energy, becomes a security concern. Dependence on external sources for these inputs risks translating into constraints on states' autonomy in the development, deployment, and sustainment of AI capabilities particularly for defensive purposes.

Starting with data, as discussed, in the development of AI, ML and DL models, data represents the fuel for AI and, consequently, the effectiveness of AI systems is constrained in a narrow way by the quality, quantity, and nature of the data they process. To achieve reliable real-world applicability, AI models require large standardized, representative, and high-quality datasets capable of properly training the systems. However, the creation of adequate dataset for training algorithms is not always straightforward (Morley & Floridi, 2025). Data scarcity and limited availability of labeled training data as one of the significant challenges facing AI development and its potential growth, especially in sectors where representative, high-quality data are essential for reliable performance (Abdalla et al., 2025; Hidary, 2025). This issue has been evident in both case studies, where inadequate datasets emerged as a key challenge. In the defense sector suitable datasets are unavailable for certain applications, due to security constraints or the lack of sufficiently specific data. Accordingly, within the healthcare sector, the field with the greatest potential for expansion is radiology, primarily because of the vast availability of data generated through standardized medical imaging formats. The risk, therefore, is the deployment of poorly trained AI systems that fail to enhance performance and may also introduce new vulnerabilities and error pathways. In recent years, the use of synthetic data has emerged as a potential response to this challenge. Synthetic data consist of artificially generated information—created through statistical methods or AI techniques—that mimics real-world data and can be used to train ML models (Caballar, n.d.). In the healthcare sector, synthetic data seem can support the acceleration of drug discovery by providing pharmaceutical companies with large volumes of data on which to train and test models, as well as medical research by providing similar real-world datasets for clinical trials (Caballar, n.d.). Even in the military domain, companies that offer synthetic data

designed to replicate real-world operational conditions are emerging. However, it will be important to assess the reliability of these synthetic data.

Closely related to data availability and use are issues of privacy and data protection. In the healthcare sector, where patient data are processed, privacy concerns are particularly acute—especially considering the rise in cyberattacks targeting medical records and digital health infrastructures (OECD, 2024*a*). In military contexts, several institutions have raised concerns regarding privacy, particularly in relation to risks to citizens' privacy and civil liberties arising from excessive or intrusive AI-enabled surveillance, as well as regarding operational security—notably the risk of leakage of classified or otherwise sensitive information through AI systems, especially in increasingly interconnected networked environments (European Defence Agency, 2025; Brescia, n.d.). With ever-increasing volume of data—often personal and, at times, sensitive—the manner in which these data are collected, managed, and potentially used, for example by AI systems, becomes a critical concern. The concept of privacy itself has evolved with the emergence of new technologies and advanced data analytics. As a result, it has become increasingly difficult for private individuals as well as institutions to exercise control over their data and to protect it from unauthorized access by third parties. The associated risks include violations of individuals privacy and sensitive information and the misuse of data for malicious purposes. As Floridi & Taddeo (2016) observe, public awareness of the value of data and the opportunities, risks and challenges associated with data science is lacking. In such an evolving scenario, robust data protection standards and effective oversight mechanisms governing how data are collected and for what purposes are essential to prevent sensitive data from being misused or concentrated in the hands of large technology companies without adequate control and accountability.

To this end, comprehensive regulatory frameworks are required to guarantee transparency and the protection of privacy in data collection, processing, and storage, as well as greater cooperation and alignment between states' policies. In this regard, Europe has been the most proactive actor in regulating data governance, most notably through the adoption of the General Data Protection Regulation (GDPR). The U.S., instead, do not have a comprehensive federal data privacy law and relies on sector specific laws and state regulations, such as California Consumer Privacy Act (CCPA) and the Health Insurance Portability and Accountability Act (HIPAA) (Bakare et al., 2024). In China data privacy is primarily governed by the Personal Information Protection Law (PIPL), a comprehensive national-level regulation which regulates the collection, use, storage

and sharing of information of individuals in China (Calzada, 2022). This fragmentation and difference in standards risk proving inadequate in addressing a phenomenon that transcends national boundaries and touches multiple jurisdictions. For instance, the practice of collecting personal data to train AI algorithms raises significant privacy concerns and requires collective and coordinated responses (European Data Protection Board, 2024). For these reasons, the OECD, in its report *AI, Data Governance And Privacy Synergies And Areas Of International Co-Operation* (2024*b*), clearly advocates for increased international cooperation and greater alignment of data protection policy frameworks, including through the establishment of multi-stakeholder communities involving public authorities, private actors, and other relevant data stakeholders.

A second critical dimension of the AI competition concerns computing power, understood in terms of data centers, semiconductors and advanced chips. The expansion of data centers and the production of cutting-edge semiconductors entail a growing dependence on a wide range of critical raw materials required for their construction and operation, including copper, steel, aluminum, gallium, germanium, and palladium (U.S. Geological Survey, 2025). As a result, competition over AI capabilities increasingly extends beyond software and algorithms to encompass the control of material supply chains, which are characterized by a high degree of global interdependence.

In this context, the analysis of military AI has already highlighted the broader security implications associated with supply chain control (Section 2.3). The U.S. has sought to limit China's access to advanced semiconductors through export controls on high-end chips and manufacturing equipment, pursued in coordination with key allies that dominate critical nodes of the global semiconductor supply chain, such as the Netherlands, Japan, South Korea, and Taiwan[10] (Carchidi & Soliman, 2024; Allen & Goldston, 2025). However, China retains significant leverage over the supply chains of several critical raw materials essential for data center and semiconductor production—most notably gallium and germanium—and has responded by imposing export restrictions on these materials, while also accelerating efforts to develop supply chains independent

[10] The Netherlands plays a pivotal role as the producer of extreme ultraviolet (EUV) lithography machines, manufactured by the Dutch firm ASML, which are essential for fabricating advanced chips. Japan is also a major supplier of semiconductor manufacturing equipment and materials, while South Korea—home to Samsung—is one of the world's leading producers of advanced semiconductors. Taiwan represents the most critical hub in advanced chip manufacturing, accounting for approximately 60% of global semiconductor production and around 90% of the world's most advanced chip-manufacturing capacity (Carchidi & Soliman, 2024; International Trade Administration (U.S.), 2025). The U.S. is therefore highly dependent on TSMC's most advanced chips, which are used by U.S. firms such as Nvidia (Edwards, 2024).

of Western inputs. These dynamics illustrate how AI-related supply chains increasingly function as instruments of power and vulnerability, reinforcing patterns of strategic interdependence. For instance, for the European Union the need to ensure resilient and secure supply chains for critical materials emerges as a key strategic imperative. Therefore, it is essential that these dynamics do not translate into structural dependencies that undermine states' capacity to develop, deploy, and sustain AI capabilities and, consequently, broader defensive capabilities.

Finally, data centers for AI are highly energy- and water-intensive. They require substantial amounts of electricity to sustain their computational functions—accounting for approximately 1.5 percent of global electricity consumption in 2024 (International Energy Agency (IEA), n.d.)—as well as large volumes of water for cooling purposes, with some facilities reportedly consuming up to 1.9 million liters of water (Turner Lee & West, 2025). As a result, the expansion of AI infrastructure is also associated with a high demand of energy and water, which raise a further question regarding their environmental and climate impact. At the moment, most of the electricity used by data centers is derived from fossil fuels, although is increasing the option of renewable energy and nuclear power (International Energy Agency (IEA), n.d.).

The present analysis highlights how, for states and regions lacking access to these critical resources, the development of mechanisms to ensure that such dependence does not become a vulnerability or a source of inequality becomes a highly important security objective. In this context, stable international partnerships become essential to avoid falling behind technologically and to prevent the emergence of structural dependencies that could undermine national and regional security. In the absence of such cooperation, there is a risk that only a limited group of actors will be able to mobilize the necessary resources, thereby generating significant global imbalances in capabilities among states.

4.3 The Role of the Private Sector in AI Development

Another critical risk that emerges from this analysis concerns the erosion of state control over the sector driven by the central role of private actors in AI development. Indeed, a key dimension of this discourse is the role of the private companies in AI development and their relationship with public institutions. As the UN Conference on Trade and Development

(UNCTAD, 2025) highlights, while the private sector has historically driven technological innovation, the current level of understanding and control over AI is unparalleled. Major technology firms have emerged as the primary researchers and developers of AI systems and, due to the concentration of market power, can be described as an oligopolistic structure. Illustrative examples include Alphabet's acquisition of the UK-based DeepMind in 2014 and Microsoft's strategic partnership with OpenAI in 2019 (UNCTAD, 2025). This central role also has conferred them a distinct foothold in the competition for AI dominance. Indeed, in the global race for AI dominance, American firms—such as Alphabet (Google), Amazon, Apple, Meta, and Microsoft—and their Chinese counterparts, including Baidu, Alibaba, and Tencent, have assumed a central role within their respective national AI strategies (Zhang et al., 2025).

In this regard, the U.S. provides an illustrative case, given its strictly market-driven economy, whereas in China is less present due to strong central government control. In the U.S. context, the intensifying strategic competition with China has increasingly framed AI as a matter of national security, with China portrayed as a technological threat to U.S. strategic interests (Sayler, 2020). This securitization of AI has, in turn, reinforced the influence of Big Tech by positioning technological advancement and leadership in AI as a top national priority (Zhang et al., 2025). Indeed, between 2016 and 2019, growing concerns regarding the expanding dominance of large technology firms and their extensive societal influence—exemplified by the Cambridge Analytica case—prompted calls for stronger regulatory oversight. However, the strategic competition with China limited the scope for such restrictive regulations, as technology companies were perceived as essential actors in maintaining U.S. technological superiority (Zhang et al., 2025). As a consequence, Big Tech firms, for a time, have been able to avoid extensive regulation due to their central role in the AI race. Only more recently, particularly following the release of ChatGPT—widely perceived as symbolizing U.S. leadership in generative AI—the U.S. government has sought to strike a balance between preserving AI supremacy and addressing domestic governance concerns (Zhang et al., 2025). This shift is reflected in the 2023 Executive Order *Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (EO 14110). Nevertheless, Big Tech companies continue to play a prominent role in the AI sector.

A report by Mélanie Gornet and Winston Maxwell, *The European approach to regulating AI through technical standards* (2024) published in the Internet Policy Review, also suggests concerns relating to the EU AI Act with respect to the potential influence of private actors in its

governance of AI. The Act relies on a conformity assessment mechanism based on harmonised European standards (hENs), compliance that can be demonstrated through certification by recognized European Standardisation Organisations—namely CEN, CENELEC, and ETSI (European Commission, n.d.*a*). These organizations, however, operate as private or semi-private bodies. While such mechanisms have long been used to regulate technical safety standards, their application within the AI Act is especially sensitive, as the regulation extends beyond technical security requirements to encompass fundamental rights protection and human security considerations. Gornet and Maxwell highlight that entrusting private standardization bodies with the implementation of fundamental rights requirements raises concerns about democratic legitimacy, accountability, and the effective protection of those rights. These mechanisms risk placing excessive reliance on private governance mechanisms risks shifting normative authority away from public institutions, potentially weakening safeguards for fundamental rights (see also Berendt et al., 2025; Corporate Europe Observatory, 2025).

Therefore, a delicate governance landscape emerges, shaped by the considerable influence exercised by private companies in the AI sector—an influence derived both from their central role in AI development and from the high level of technical expertise required to design and deploy these systems. This situation is further complicated by the intensifying strategic rivalry between the U.S. and China in the field of AI, which makes coordinated efforts to constrain the power of large technology firms even more challenging. The consequently rapid pace of deployment of technological innovation is placing traditional bureaucratic systems and institutional structures under strain. Beyond the case of Elon Musk's Starlink satellite network (discussed in Section 2.3), the growing presence of cutting-edge technologies originating in the private sector is pushing military institutions to adopt approaches more closely aligned to Silicon Valley mindset, characterized by rapid action, a higher tolerance for risk, and the acceptance or adaptation to failure (Csernatoni et al., 2025). As a result, contemporary battlefields—such as the ongoing conflict in Ukraine—have increasingly become laboratories for the testing, adaptation, and deployment of emerging technologies (Csernatoni et al., 2025). However, this growing dependence on private actors raises significant concerns regarding governance, accountability, and regulatory oversight. The uncertainty surrounding the development of AI technologies, together with the potential for unforeseen and destabilizing effects, suggests a gradual and regulated integration of AI into society—one that is not driven by the interests of Big Tech (Zhang et al., 2025). In this context,

major international institutions, including the UN and the European Union, have emphasized the importance of multi-stakeholder cooperation involving private companies, research institutions, sectoral experts, and legal scholars in order to promote AI development that is responsible, ethical, accessible, and inclusive (UNCTAD, 2025; European Commission, n.d.*b*). To avoid excessive concentration of power in the hands of large technology firms over a technology with profound societal and security implications, it is essential to develop robust regulatory safeguards accompanied by sustained multi-stakeholder engagement.

4.4 AI Regulation and Global Governance

Finally, both the military and healthcare sectors have highlighted the need for appropriate regulations for the development and implementation of AI systems. This task appears to be highly challenging for several reasons. First, as discussed in Chapter 1, AI is inherently difficult to define in all its features and potential applications, which complicates the drafting of regulatory frameworks capable of encompassing the full range of existing and future technological uses. Second, the rapid evolution of AI technologies poses significant challenges for legislators, who often struggle to keep pace with ongoing developments. The European AI Act, for instance, took a long amount of time to draft also because the technology continued to evolve in the meantime, raising the risks that the regulation could become outdated by the time of its adoption. Additionally, the versatility of AI across multiple sectors necessitates the development of sector-specific regulations. With applications ranging from virtual nurses to LAWS, a single regulatory framework is insufficient. This diversity of uses further complicates the AI governance landscape. Finally, regulatory frameworks must be designed in a way that ensures the ethical and sustainable development of AI while avoiding unnecessary constraints on innovation. A delicate balance must be achieved between promoting technological advancement and ensuring regulatory compliance (Walter, 2024).

This is not an easy task, as evidenced by the existence of different approaches and models of AI regulation. The initiatives undertaken in recent years to govern AI reflect divergent regulatory strategies shaped by the priorities of the institutions involved. In the EU, the primary regulatory framework for AI is the EU AI Act (Regulation (EU) 2024/1689, 2024). The drafting process began in 2021, following a proposal by the European Commission, and the Act was formally adopted in 2024 (EU Artificial Intelligence Act, n.d.). The EU AI Act adopts a comprehensive, risk-based

approach, categorizing AI systems into four levels of risk: (i) *unacceptable risks*: systems posing a serious threat to individuals, such as applications involving cognitive behavioral manipulation or social scoring, and that for this reason are banned; (ii) *high risk*: systems affecting health, safety or fundamental rights—like applications in critical infrastructures (e.g. transport) or components of products (like robot-assisted surgery)—and are subject to stringent requirements, including rigorous risk assessment; (iii) *limited risk*: systems that present relatively low risk of negatively impact society, like AI chatbots, and have minimal transparency obligations, to guarantee that users are informed when they interact with AI; (iv) *minimal risks*: systems that pose the lowest risk and are not subject to specific obligations under the AI Act. Moreover, the EU AI Act further distinguishes AI systems into three categories—General Purpose AI, Foundation Models, and Generative AI—based on their functionality and potential impact (Regulation (EU) 2024/1689, 2024). Through this framework, the EU aimed to establish a comprehensive legal regime for AI governance—the first of its kind in the world.

Regarding the U.S., in 2023 former U.S. President Joe Biden signed the *Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence* (EO 14110). The order marked a pivotal moment in the U.S.'s effort to manage the risks and opportunities of AI. It contained a set of actions aimed at ensuring the safety and trustworthiness of AI systems while preserving America's leading position in AI innovation by promoting research. Indeed, the U.S.'s approach to AI regulation seeks to strike a balance between ensuring responsible AI development and maintaining its role as a global technological leader—particularly in light of competitive pressures from China, its second-place competitor (Walter, 2024). To this end, industry stakeholders play a central role in the regulatory process: major technology companies and sector experts are involved in shaping regulatory discussions, for example through participation in congressional hearings. So far, no comprehensive binding legislation comparable to the EU AI Act has been introduced and sector-specific regulatory pathways have been preferred. This is partly due to the fear of establishing overly rigid legislation that could constrain innovation and undermine the dynamic development of the AI sector (Walter, 2024).

China, the world's second-largest power in the sector, has adopted a regulatory approach characterized by two primary objectives: promoting AI innovation while simultaneously ensuring state control over the technology. Indeed, China's approach to AI regulation is closely reflective of its political system and governance model. With social and political stability as a central priority, China has been one of the first introducing AI-specific regulatory frameworks. In 2023, was

published the *Interim Measures for the Management of Generative AI Services*, with the aim of mitigating a range of risks associated with the use of AI services by the public (Zou & Zhang, 2025). China's AI governance framework has been considerably shaped to prioritize security-oriented applications, although recent developments in China's regulation landscape indicate an increasing tendency to balance security interests with the desire to reduce constraints on innovation (Cheng & Zeng, 2023).

The overview presented highlights a clear divergence of approaches to regulating AI among the major powers, reflecting competing conceptions of the relationship between innovation, security, and governance. The EU has adopted a comprehensive, risk-based regulatory approach, addressing AI horizontally by framing the risks associated with this technology as a cross-sectoral issue. The framework does not explicitly cover certain key areas, such as the military use of AI, as national security continues to fall under the responsibility of Member States. However, by regulating dual-use technologies developed in the civilian sphere, it indirectly influences military applications[11] (Powell, 2024). The U.S. adopts a more decentralized approach focused on prioritizing individual innovation and market-driven development, while ensuring the safe development through specific-sector regulation. By contrast, China pursues a hybrid approach that combines state control—particularly to support surveillance and security capabilities—with innovation promotion[12].

Such fragmentation appears inadequate for addressing the systemic risks associated with AI technologies that extend beyond national borders. The issues emerged from both the sectors analyzed underscore the necessity of a coordinated global strategy [13]. The unprecedented levels of technical uncertainty, a market structure favoring regional customization over global uniformity, and intensifying tech rivalry between the U.S. and China hinder the establishment of worldwide AI regulations. The EU has attempted to set a global standard thorough the EU AI Act; however, the

[11] The EU AI Act excludes AI technologies used exclusively for national security and defense purposes, while in cases of dual-use technologies it continues to apply (Powell, 2024).
[12] As Hine & Floridi (2024) note, AI regulation in both the U.S and China also reflects deeper societal and ethical differences rooted, respectively, in Protestant ethics and Confucianism.
[13] Effects on the labor market represent another major area of concern that should also be considered within a broader security analysis. Kristalina Georgieva has warned that the International Monetary Fund (IMF) expects 60% of jobs in advanced economies to be affected by AI in the coming years (Wearden & Stewart, 2026). However, this issue is not addressed in the present thesis, as it is not directly related to the selected case studies. In the military domain, labor-market effects appear to be limited or even positive, insofar as AI applications may reduce human exposure in high-risk environments. In the healthcare sector, there is broad consensus that AI is more likely to augment rather than replace human labor, a conclusion also supported by the applications analyzed in this thesis, which point toward models of human–machine collaboration rather than fully automated systems.

Brussels Effect has had limited impact in this domain, as AI systems, unlike other products, can easily be tailored to specific jurisdictions (Crum, 2025). Nevertheless, given the transnational and dual-use nature of AI, coordinated global action is needed to prevent governance gaps that may generate systemic instability and to ensure that AI functions as a driver of greater security rather than insecurity and fragmentation. For this reason, a collective awareness and coordinated effort are required to develop global guidance aimed at ensuring the safe and sustainable development of AI. A positive signal has been the growing number of international forums addressing this issue—such as the G7 in 2024, during which states affirmed their commitment to the development of AI that is safe, secure, and trustworthy, as well as the AI Action Summit, which since 2023 has brought together government representatives to discuss pathways for the responsible development of AI. It is important to continue advancing along this path in order to establish shared global governance practices for the responsible and safe use of AI.

Conclusions

What emerges from this analysis is a highly complex and interconnected landscape, characterized by multiple overlapping dynamics in an already challenging geopolitical environment. The examination of the two sectors reveals a dual dynamic. On the one hand, AI offers significant possibilities and enhancement in both domains. Its data-processing capabilities offer substantial opportunities by enhancing core functions such as data collection, advanced data analysis, pattern recognition, and the prediction of possible courses of action. The analysis of the military domain and healthcare has shown how these capabilities can be applied in such diverse fields. In the military domain, AI applications contribute to improved intelligence analysis, enhanced interoperability among forces, and increased speed and precision in responses, thereby strengthening defensive capabilities and enabling a more interconnected, prepared, and resilient security apparatus. Moreover, in a security environment characterized by the emergence of unconventional hybrid threats, AI is proving to be central also to ensuring defensive capabilities. In the healthcare sector, although applications differ substantially, AI has similarly demonstrated strong potential to enhance sectoral capabilities and address some of the structural challenges currently facing the sector—such as rising demand for healthcare services and shortages of healthcare professionals—by improving system performance, optimizing resource allocation, and fostering technological advancement. Taken together, the case studies highlight how AI can make positive contributions to security in a broad sense, enhancing preparedness and resilience at multiple levels. When considering that these dynamics extend beyond the two analyzed cases and apply to a broad range of other sectors affected by AI-driven transformation—like finance, industry, and diplomacy—the picture becomes even more multifaceted.

On the other hand, the analysis also emphasized the presence of substantial challenges that accompany AI integration across both sectors. These include limited transparency and explainability of algorithm systems, risk of loss of control, gaps in accountability and responsibility across the AI lifecycle, data governance and privacy risks, the growing influence of private actors, and the absence of fully adequate regulatory frameworks. These challenges underscore potentially disruptive effects of AI implementation and call for special attention. In the military domain, such risks include potential violations of International Humanitarian Law and the loss of meaningful control over decision-making processes. In healthcare, they manifest primarily in risk of harm to individuals. Both case-studies converge on the need for adequate robust safeguards, particularly in terms of human oversight, governance, and accountability. Indeed, AI simultaneously

strengthens state capacities and introduces systemic vulnerabilities that cut across domains. Consequently, both new security opportunities and new threats emerge, and the deployment of AI increasingly depends not only on technological adoption but also on effective governance, resource control, and institutional resilience.

In this regard, the analysis highlighted several recommendations needed to mitigate the challenges. First, the establishment of robust mechanisms for testing, validation, and continuous evaluation are necessary to ensure the reliability, safety, and security of AI systems across their entire lifecycle. Second, the maintaining of meaningful human involvement is crucial to ensure responsible, controlled, and ethically grounded use of AI. Given the inherent opacity of many AI applications, informed human oversight is required—across both military and healthcare contexts—to supervise AI outputs and preserve human judgment. Third, the need for adequate education and training, aimed at preparing professionals and society to interact competently with increasingly pervasive AI technologies. Fourth, inter-sectoral dialogue and collaboration between AI developers, domain experts, and regulatory authorities is necessary to ensure the optimal development of AI systems that are context-aware, operationally appropriate, and in line with safety and ethical standards.

At the governance level, regulatory initiatives are emerging from states and international organizations—including the European Union, the U.S. Department of Defense, NATO, and the OECD—that provide important guidelines for AI development and use. However, there is still a lack of commonly accepted international standards. Given the transnational nature of AI systems, data flows, and technological infrastructures, such common frameworks appear increasingly necessary to ensure regulatory coherence and to avoid the emergence of governance gaps that may be exploited to circumvent safeguards. The development of a shared regulatory framework is hindered by the current competitive international environment, in which major technological powers often prioritize speed and strategic advantage over regulatory caution. Nevertheless, the risk presented throughout the analyses highlight the need of common efforts to ensure a responsible, safe and ethical implementation of AI technologies. Global recognition of the risks associated with uncontrolled AI development—and of the need for coordinated governance—appears essential. The tension between rapid innovation and regulatory restraint must therefore be reconciled through a balanced approach that integrates technological innovation with ethical and responsible development. The point is not to restrict innovation in the field of AI, but rather to

acknowledge the potential risks and adopt a more informed approach. This can be achieved through sustained multi-stakeholder dialogue involving governments, technology companies, sectoral experts, and legal scholars. The risk of not adopting a global governance on AI development is it will become a source of instability rather than security, exacerbating global inequalities and reinforcing structural dependencies, with only a limited group of actors able to mobilize the necessary data, capital, and energy resources. Instead, the potential of AI must strengthen collective security and institutional resilience. Greater international cooperation and regulatory alignment are needed to ensure that AI contributes to stability, security, and a sustainable global order.

Bibliography and Sitography

Abdalla, H. B., Kumar, Y., Marchena, J., Guzman, S., Awlla, A., Gheisari, M., & Cheraghy, M. (2025). *The Future of Artificial Intelligence in the Face of Data Scarcity*. Computers, Materials & Continu*a*, 84(1).

Air Force Material Command. (2025, March 27). *Advanced Battle Management System: victory through distributed connectivity.* Air Force Material Command. https://www.afmc.af.mil/News/Article-Display/Article/4137120/advanced-battle-management-system-victory-through-distributed-connectivity/ (Accessed October 23, 2025).

Akhavein, D., Sheel, M., & Abimbola, S. (2025). *Health security—Why is 'public health' not enough?.* Global Health Research and Policy, 10(1), 1.

Al Kuwaiti, A., Nazer, K., Al-Reedy, A., Al-Shehri, S., Al-Muhanna, A., Subbarayalu, A. V., Al Muhanna, D., & Al-Muhanna, F. A. (2023). *A review of the role of artificial intelligence in healthcare*. Journal of personalized medicine, *13*(6), 951.

Alderman, R. (2021, December 23). *How Rainmaker, Prome. theus, FIRESTORM, and SHOT AI algorithms enable the Kill Web*. Military embedded systems. https://militaryembedded.com/radar-ew/sensors/how-rainmaker-prometheus-firestorm-and-shot-ai-algorithms-enable-the-kill-web#:~:text=reconnaissance,using%20110%20new%20technologies%20and (Accessed July 30, 2025).

Allen, G.C., & Goldston, I. (2025, March 14). *Understanding U.S. Allies' Current Legal Authority to Implement AI and Semiconductor Export Controls*. CSIS. https://www.csis.org/analysis/understanding-us-allies-current-legal-authority-implement-ai-and-semiconductor-export.

Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). *Machine bias: There's software used across the country to predict future criminals. And it's biased against Blacks*. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing (Accessed 23 June 2025).

Archivio Disarmo (IRIAD). (2025). *Lo stato dell'Intelligenza Artificiale in ambito militare e le prospettive di regolazione a livello nazionale, europeo e internazionale*. Archivio Disarmo. https://www.esteri.it/wp-content/uploads/2025/10/Rapporto-MAECI_con-copertina_compressed.pdf.

Autonomous Weapons (2025, April 25). *The Political Landscape: How Nations are Responding to Autonomous Weapons in War.* Autonomous Weapons. https://autonomousweapons.org/global-perspectives-on-regulation/#:~:text=According%20to%20Automated%20Decision%20Research,54%20(28%25)%20remaining%20undecided. (Accessed July 8, 2025).

Bakare, S. S., Adeniyi, A. O., Akpuokwe, C. U., & Eneh, N. E. (2024). *Data privacy laws and compliance: a comparative review of the EU GDPR and USA regulations*. Computer Science & IT Research Journal.

Baskaran, G., & Schwartz, G. (2024, October 7). *From Mine to Microchip. Addressing Critical Mineral Supply Chain Risks in Semiconductor Production*. CSIS. https://www.csis.org/analysis/mine-microchip.

Berendt, B., Danos, V., Hartmann, D., Langer, F., Lassiter, T., Mönig, J. M., Mysegades, C., Puntschuh, M., & Zech, H. (2025). *Harmonised Standards and Conformity Assessments in the AI Act: Strengthening Independent and Participatory Oversight*. Weizenbaum Institute. https://www.weizenbaum-institut.de/media/Publikationen/Weizenbaum_Policy_Paper/Weizenbaum_Policy_Paper_17.pdf.

Bondar, K. (2024, December 11). *Does Ukraine Already Have Functional CJADC2 Technology?*. Centre for Strategic and International Studies. https://www.csis.org/analysis/does-ukraine-already-have-functional-cjadc2-technology. (Accessed October 24, 2025).

Borchert, H., Schütz, T., & Verbovszky, J. (2024). *Master and Servant: Defense AI in Germany*. In Borchert, H., Schütz, T., & Verbovszky, J. (Eds). (2024). *The Very Long Game. 25 Case Studies on the Global State of Defense AI*. Springer.

Borchert, H., Brandlhuber, C., Brandstetter, A., & Schaal, G. S. (2022). *Free Jazz on the Battlefield. How GhostPlay's AI Approach Enhances Air Defense*. Defense AI Observatory. https://defenseai.eu/daio_study2203.

Borchert, H., Schutz, T., & Verbovszky, J. (2021). *Beware of the Hype. What Military Conflicts in Ukraine, Syria, Lybia, and Nagorno-Karabakh (Don't) Tell Us about the Future of War*. Defence AI Observatory. Helmut Schmidt University.

Brescia, S. (n.d.). *Navigating the AI battlefield: Opportunities and ethical frontiers*. NATO Headquarters Rapid Deployable Corps Italy (NRDC-ITA). https://nrdc-ita.nato.int/newsroom/insights/navigating-the-ai-battlefield-opportunities--challenges--and-ethical-frontiers-in-modern-warfare. (Accessed February 5, 2026).

Caballar, R.D. (n.d.) *What is synthetic data?*. IBM. https://www.ibm.com/think/topics/synthetic-data. (Accessed January 30, 2026).

Caballar, R.D., & Stryker, C. (n.d.). *What is open-source AI?*. IBM. https://www.ibm.com/think/topics/open-source-ai. (Accessed February 2, 2026).

Calzada, I. (2022). *Citizens' data privacy in China: The state of the art of the Personal Information Protection Law (PIPL)*. Smart Cities, 5(3), 1129-1150.

Carchidi, V., & Soliman, M. (2024). *The role of the Middle East in the US-China race to AI supremacy.* Middle East Institute.

Cecchini, A., Comparato, M., D'Ambrosio, P., Americo, M., Ardizzone, F., Baldi, P., Castellani, L., D'Ippolito, M., Guerriera, M., Polito, R., Sciò, D., Villani, R., Abdullahi, B., & Linares, V., (2023). *Applicazione di Artificial Intelligence per fini militari: individuazione dei criteri relativi al passaggio dall'approccio Human in the Loop allo Human on the Loop e definizione delle conseguenti implicazioni sul ciclo di definizione e approvazione delle ROE, con considerazioni sull'adeguamento del quadro normativo in caso di incidenti/eventi avversi.* Istituto Superiore di Stato Maggiore Interforze.

Center for Strategic and International Studies. (2025, May 28). *Russia and Ukraine: The drone war—Innovation on the frontlines and beyond*. CSIS. https://www.csis.org/analysis/russia-ukraine-drone-war-innovation-frontlines-and-beyond (Accessed July 4, 2025).

Cheng J, & Zeng J. *Shaping AI's future? China in global AI governance*. J Contemp China. 2023;32(143):794–810.

Clapp, S. (2025*a*). *Military drone systems in the EU and global context: Types, capabilities and regulatory frameworks*. European Parliamentary Research Service.

Clapp, S. (2025*b*). *Defence and artificial intelligence*. European Parliamentary Research Service.

Congressional Research Service (2022*a*, January 21). *Joint All-Domain Command and Control (JADC2)*. Congressional Research Service.

Congressional research Service (2022*b*, February 15). *Advanced Battle Management System (ABMS).* Congressional Research Service.

Corporate Europe Observatory. (2025). *How Big Tech sets its own AI standards*. Corporate Europe Observatory. https://corporateeurope.org/en/2025/01/bias-baked.

Corrigan, C.C. (2024). *AI For Human Security – Applications and Ethical Considerations*. IEAI. https://www.ieai.sot.tum.de/wp-content/uploads/2024/02/IEAIResearchBrief_Q12024_AI-for-Human-Security.pdf.

Crum, B. (2025). *Brussels effect or experimentalism? The EU AI Act and global standard-setting*. Internet Policy Review, 14(3).

Csernatoni, R., Broeders, D., Andersen, L. H., Hoijtink, M., Bode, I., Lindsay, J. R., & Schwarz, E. (2025). *Myth, Power, and Agency: Rethinking Artificial Intelligence, Geopolitics and War.* Minds and Machines, 35(3), 37.

Dahlmann, A. (2022). *Drones and Lethal Autonomous Weapon Systems*. In Reinhold, T., & Schörnig, N. (Eds.). (2022). *Armament, Arms Control and Artifcial Intelligence. The Janus-faced Nature of Machine Learning in the Military Realm.* Springer.

Davis, Z. (2019). *Artificial intelligence on the battlefield*. Prism, 8(2), 114-131. https://cgsr.llnl.gov/sites/cgsr/files/2024-08/CGSR-AI_BattlefieldWEB.pdf.

Defence Innovation Board. (2019). *AI Principles: Recommendations on the Ethical Use of Artificial Intelligence by the Department of Defence*. Defence Innovation Board.

Dembrower, K., Crippa, A., Colón, E., Eklund, M., & Strand, F. (2023). *Artificial intelligence for breast cancer detection in screening mammography in Sweden: a prospective, population-based, paired-reader, non-inferiority study*. The Lancet Digital Health, 5(10), e703-e711.

Dina, A. (2024, September 19). *Palantir Maven Smart System (MSS) contratto $100M con U.S. Army.* Rivista AI. https://www.rivista.ai/2024/09/19/palantir-maven-smart-system-mss-contratto-100m-con-u-s-army/ (Accessed July 16, 2025).

Dixon, D., Sattar, H., Moros, N., Kesireddy, S. R., Ahsan, H., Lakkimsetti, M., Fatima, M., Doshi, D., Sadhu, K., & Hassan, M., J. (2024). *Unveiling the influence of AI predictive analytics on patient outcomes: a comprehensive narrative review*. Cureus, 16(5).

Dolinko, I., & Antebi, L. (2024). *Embracing the Organized Mess: Defense AI in Israel*. In Borchert, H., Schütz, T., & Verbovszky, J. (Eds). (2024). *The Very Long Game. 25 Case Studies on the Global State of Defense AI.* Springer.

Dumbacher, E. (2025, December 29). *How Deepfakes Could Lead to Doomsday*. Foreign Affairs. (Accessed January 11, 2026)

Edwards, J. (2024). *Chips, subsidies and commercial competition between the United States and China.* In Song, L., & Zhou, Y. (Eds). *China: Regaining Growth Momentum After the Pandemic*. ANU Press.

EU Artificial Intelligence Act (n.d.). *Historic Timeline*. https://artificialintelligenceact.eu/developments/ (Accessed January 24, 2026)

European Centre of Excellence for Countering Hybrid Threats (Hybrid CoE). (n.d.). *Deterrence and Resilience*. Hybrid CoE. https://www.hybridcoe.fi/deterrence-and-resilience/. (Accessed February 13, 2026).

European Commission: Directorate-General for Health and Food Safety, EEIG, Open Evidence & PwC. (2025). *Study on the deployment of AI in healthcare: final report*. Publications Office of the European Union. https://data.europa.eu/doi/10.2875/2169577.

European Commission. (2021, June 30). *AI4DEF*. European Commission. https://defence-industry-space.ec.europa.eu/ai4def_en. (Accessed November3, 2025).

European Commission. (n.d.*a*). *Harmonised Standards*. European Commission. https://single-market-economy.ec.europa.eu/single-market/goods/european-standards/harmonised-standards_en. (Accessed February 4, 2026).

European Commission. (n.d.*b*). *Apply AI Alliance*. European Commission. https://digital-strategy.ec.europa.eu/en/policies/apply-ai-alliance. (Accessed February 4, 2026).

European Commission. (n.d.*c*). *Cybersecurity of hospitals and healthcare providers*. European Commission. https://commission.europa.eu/topics/digital-economy-and-society/cybersecurity-healthcare_en. (Accessed February 4, 2026).

European Commission. High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. European Commission. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

European Data Protection Board. (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*. https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf.

European Defence Agency. (2020). *European Defence Matters: Enhancing interoperability – Train together, deploy together*. https://eda.europa.eu/docs/default-source/eda-magazine/edm19_web.pdf.

European Defence Agency. (2025). *WHITEPAPER: Trustworthiness for Artificial Intelligence in Defence*. European Defence Agency. https://eda.europa.eu/docs/default-source/brochures/taid-white-paper-final-09052025.pdf.

Ewing, J. (2024, October 10). *Elon Musk teases Tesla's robotaxi, again*. The New York Times. https://www.nytimes.com/2024/10/10/business/tesla-robotaxi-elon-musk.html (Accessed 23 June 2025).

Executive Office of the President. National Science and Technology Council Committee on Technology (2016). *Preparing for the future artificial intelligence*. Executive Office of the President. https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf.

Farrar, O. (2025). *Understanding AI Vulnerabilities. As artificial intelligence capabilities evolve, so too will the tactics used to exploit them*. Harvard Magazine.

Food and Drug Administration (FDA) (U.S.). (n.d.). *Artificial Intelligence-Enabled Medical Devices*. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices#:~:text=To%20support%20transparency%20in%20the,arrows%20in%20the%20column%20headings. (Accessed December 26, 2025)

Fedasiuk, R., & Weinstein, E. (2022). *AI in the Chinese military*. In Hannas, W.C., & Chang, H.M. (Eds) (2022). *Chinese power and artificial intelligence. Perspectives and Challenges*. Routledge.

Fischer, S. (2022). *Military AI applications: A cross-country comparison of emerging capabilities in arms control for artificial intelligence*. In Reinhold, T., & Schörnig, N. (Eds.), *Armament, arms control and artificial intelligence: The Janus-faced nature of machine learning in the military realm*. Springer

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). *Delivering Cognitive Behavior Therapy to Young Adults With Symptoms of Depression and Anxiety Using a Fully Automated Conversational Agent (Woebot): A Randomized Controlled Trial*. JMIR Mental Health, 4(2), e7785.

Floridi, L. & Taddeo, M. (2016). *What is data ethics?* Phil. Trans. R. Soc. A 374:20160360.

Gao, X., He, P., Zhou, Y., & Qin, X. (2024). *Artificial intelligence applications in smart healthcare: a survey*. Future Internet, 16(9), 308.

Gornet, M., & Maxwell, W. (2024). *The European approach to regulating AI through technical standards*. Internet Policy Review. https://policyreview.info/pdf/policyreview-2024-3-1784.pdf.

Graae, A. I. (2024). *Servers Before Tanks? Defence AI in Denmark.* In Borchert, H., Schütz, T., & Verbovszky, J. (Eds). *The Very Long Game. 25 Case Studies on the Global State of Defense AI.* Springer.

Graham, T., & Singer, P.W. (2025, March 2). *New products show China's quest to automate battle*. Defense One. https://www.defenseone.com/threats/2025/03/new-products-show-chinas-quest-automate-battle/403387/#:~:text=All%20this%20is%20part%20of,code%2C%20and%20accelerate%20weapons%20development (Accessed August 28, 2025).

Grand-Clément, S. (2023). *Artificial Intelligence beyond weapons: Application and impact of AI in the military domain*. UNIDIR. Geneva.

Gudigar, A., Raghavendra, U., Nayak, S., Ooi, C.P., Chan, W.Y., Gangavarapu ,M.R., Dharmik, C., Samanth, J., Kadri N.A., & Hasikin, K. (2021). *Role of artificial intelligence in COVID-19 detection*. Sensors, 21(23), 8045.

Gutlove, P., & Thompson, G. (2003). *Human security: Expanding the scope of public health*. Medicine, Conflict and Survival, 19(1), 17-34.

Hamilton, A. (2024). *Artificial intelligence and healthcare simulation: the shifting landscape of medical education*. Cureus, 16(5).

Hidary, J. (2025, December 6). *We're running low on data to train AI. The good news is there's a fix for that.* World Economic Forum. https://www.weforum.org/stories/2025/12/data-ai-training-synthetic/ (Accessed January 18, 2026).

Hine, E., & Floridi, L. (2024). *Artificial intelligence with American values and Chinese characteristics: a comparative analysis of American and Chinese governmental AI policies*. Ai & society, 39(1), 257-278.

Ho, C., Tsakonas, E., Tran, K., Cimon, K., Severn, M., Mierzwinski-Urban, M., & Corcos, J. (2012). *Robot-assisted surgery compared with open surgery and laparoscopic surgery.* CADTH technology overviews, 2(2), e2203.

Hodges, B. (2025, July 8). *Air Force advances human-machine teaming with autonomous collaborative platforms*. Air Force Research Laboratory Public Affairs. https://www.af.mil/News/Article-Display/Article/4236861/air-force-advances-human-machine-teaming-with-autonomous-collaborative-platforms/. (Accessed February 5, 2026).

Holdsworth, J., & Kosinski, M. (n.d.). *What is a distributed denial-of-service (DDoS) attack?*. IBM. https://www.ibm.com/think/topics/ddos. (Accessed February 12, 2026).

Holmgren, A. J., Sinsky, C. A., Rotenstein, L., & Apathy, N. C. (2024). *National comparison of ambulatory physician electronic health record use across specialties*. Journal of general internal medicine, 39(14).

Horowitz, M. C. (2018). *Artificial intelligence, international competition, and the balance of power*. Texas National Security Review: Volume 1, Issue 3.

Horowitz, M. C., Kania, E. B., Allen, G. C., & Scharre, P. (2018). *Strategic competition in an Era of artificial intelligence*. Center for a New American Security. https://www.cnas.org/publications/reports/strategic-competition-in-an-era-of-artificial-intelligence.

Hudec, M., & Smutny, Z. (2017). *RUDO: A home ambient intelligence system for blind people.* Sensors, *17*(8), 1926.

Hunter, L.Y., Albert, C.D., Rutland, J., Topping, K., & Hennigan, C. (2024). *Artificial intelligence and information warfare in major power states: how the US, China, and Russia are using artificial intelligence in their information warfare and influence operations.* Defense & Security Analysis. *40*(2), 235-269.

Ibrahim, A. (2024, December 15). *United States' Project Maven And The Rise Of AI-Assisted Warfare*. Global Defense Insight. https://defensetalks.com/united-states-project-maven-and-the-rise-of-ai-assisted-warfare/ (Accessed July 16, 2025).

International Energy Agency (IEA). (n.d.). *Energy supply for AI*. International Energy Agency. https://www.iea.org/reports/energy-and-ai/energy-supply-for-ai. (Accessed February 4, 2026).

International Energy Agency. (n.d.). *Energy demand from AI*. International Energy Agency. https://www.iea.org/reports/energy-and-ai/energy-demand-from-ai. (Accessed January 27, 2026).

International Trade Administration (U.S.). (2025, December 1). Taiwan - Semiconductors including chip design for AI. U.S. Department of Commerce. https://www.trade.gov/country-commercial-guides/taiwan-semiconductors-including-chip-design-ai. (Accessed January 26, 2026).

Jensen, B., & Paschkewitz, J. (2019, December 23). *Mosaic Warfare: Small and Scalable are Beautiful.* War on the rocks. https://warontherocks.com/2019/12/mosaic-warfare-small-and-scalable-are-beautiful/. (Accessed September 1, 2025).

Johnson, J. (2021). *Artificial intelligence and the future of warfare: The USA, China, and strategic stability*. Manchester University Press.

Kahn, L. A. (2024). *Risky Incrementalism: Defense AI in the United States*. In Borchert, H., Schütz, T., & Verbovszky, J. (Eds). *The Very Long Game. 25 Case Studies on the Global State of Defense AI.* Springer.

Khandalavala, K., Shimon. T., Flores, L., Armijo, P.R., & Oleynikov, D. (2020). *Emerging surgical robotic technology: A progression toward Microbots*. Ann laparoscendosc surg. 5.

Kim, J., Campbell, A. S., & Wang, J. (2018). *Wearable non-invasive epidermal glucose sensors: A review.* Talanta, 177, 163-170.

Kong, A., & Shaoshan, L. (2022, July 1). *How autonomous mobile clinics can transform healthcare in least developed countries*. World Economic Forum. https://www.weforum.org/stories/2022/07/autonomous-mobile-clinics-healthcare/ (Accessed January 6, 2026).

Kosinski, M. (n.d.). *What is phishing?*. IBM. https://www.ibm.com/think/topics/phishing. (Accessed January 6, 2026).

Kühl, N., Schemmer, M., Goutier, M., & Satzger, G. (2022). *Artificial intelligence and machine learning*. Electronic Markets.

Lacdan, J. (2021, August 16). *Project Convergence 21 to showcase abilities of the joint force. U.S. Army*. https://www.army.mil/article/249422/project_convergence_21_to_showcase_abilities_of_the_joint_force. (Accessed February 17, 2026).

Leonardo. (2022, October 17). *Il Security Operation Centre di Chieti: un sistema di competenze digitali contro il pericolo cyber.* Leonardo. https://www.leonardo.com/it/news-and-stories-detail/-/detail/leonardo-soc-chieti-cyberthreats. (Accessed February 12, 2026).

Liu, N., Zhang, Z., Ho, A. F. W., & Ong, M. E. H. (2018). *Artificial intelligence in emergency medicine.* Journal of Emergency and Critical Care Medicine, 2.

Lonergan, E. (2025, November 24). *Exploring the Implications of Military Artificial Intelligence for Deterrence*. Perry World House.

Malik, A., & Solaiman, B. (2024). *AI in hospital administration and management: Ethical and legal implications.* Research Handbook on Health, AI and the Law, 20-40.

Mallick, P. K., & Gen, M. (2019). *Is Artificial Intelligence (AI) Changing the Nature of War*. Vivekananda International Foundation, 18. https://www.vifindia.org/article/2019/january/18/is-artificial-intelligence-changing-the-nature-of-war (Accessed July 5, 2025).

Martin, K., & Liversain, L. (2024). *A Winding Road Before Scaling-Up? Defense AI in France*. In Borchert, H., Schütz, T., & Verbovszky, J. (Eds). *The Very Long Game. 25 Case Studies on the Global State of Defense AI.* Springer.

Maxwell, P. (2020, April 20). *Artificial Intelligence Is the Future of Warfare (Just Not in the Way You Think). Modern.* War Institute at West Point. https://mwi.westpoint.edu/artificial-intelligence-future-warfare-just-not-way-think/ (Accessed July 7, 2025).

McCoy, D., Roberts, S., Daoudi, S., & Kennedy, J. (2023*). Global health security and the health-security nexus: principles, politics and praxis*. BMJ Global Health, 8(9), e013067.

Mimran, T., & Dahan, G. (2024, April 20). *Artificial Intelligence in the Battlefield: A Perspective from Israel*. OpinioJuris. https://opiniojuris.org/2024/04/20/artificial-intelligence-in-the-battlefield-a-perspective-from-israel/#:~:text=Recently%2C%20high%2Dranking%20officers%20in,for%20defensive%20needs%2C%20command%20and (Accessed September 2, 2025).

Misuraca, G., & Van Noordt, C. (2020). *AI Watch - Artificial Intelligence in public services*. Publications Office of the European Union. EUR 30255 EN. Luxembourg.

Morley, J., & Floridi, L. (2025). *The ethics of AI in healthcare: An updated mapping review*. Ethics and Medical Technology: Essays on Artificial Intelligence, Enhancement, Privacy, and Justice, 29-57.

Mudgal, S. K., Agarwal, R., Chaturvedi, J., Gaur, R., & Ranjan, N. (2022). *Real-world application, challenges and implication of artificial intelligence in healthcare: an essay*. Pan African Medical Journal, *43*(1).

Natarajan, A., Su, H. W., & Heneghan, C. (2020). *Assessment of physiological signs associated with COVID-19 measured using wearable devices*. NPJ digital medicine, 3(1), 156.

National Gaucher Foundation (n.d.). *Using Artificial Intelligence to Diagnose Rare Genetic Diseases*. https://www.gaucherdisease.org/blog/ai-and-rare-disease-diagnosis-national-gaucher-foundation/ (Accessed November 16, 2025).

National Security Commission on Artificial Intelligence (NSCAI) (U.S.). (2021). *Final Report*. National Security Commission on Artificial Intelligence. https://assets.foleon.com/eu-central-1/de-uploads-7e3kk3/48187/nscai_full_report_digital.04d6b124173c.pdf.

NATO Parliamentary Assembly. (2025). *Protecting Allied Societies from Disinformation Emanating from the People's Republic of China*. NATO Parliamentary Assembly. Report No. 011 CDSRCS 25 E, rev.2. https://www.nato-pa.int/download-file?filename=/sites/default/files/2025-10/011%20CDSRCS%2025%20E%20rev.2%20fin%20-%20DISINFORMATION%20EMANATING%20FROM%20THE%20PRC%20-%20TEITELBAUM%20REPORT.pdf.

NATO. (2023*a*, March 22). *Electromagnetic warfare*. https://www.nato.int/cps/en/natohq/topics_80906.htm. (Accessed October 28, 2025).

NATO. (2023*b*). *Alliance Persistent Surveillance from Space (APSS).* https://www.nato.int/nato_static_fl2014/assets/pdf/2023/2/pdf/230215-factsheet-apss.pdf (Accessed September 3, 2025).

NATO. (2024, July 10). *Summary of NATO's revised Artificial Intelligence (AI) strategy*. NATO. https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy. (Accessed January 31, 2026)

NATO. (2025, April 14). *NATO acquires AI-enabled Warfighting System*. https://shape.nato.int/news-releases/nato-acquires-aienabled-warfighting-system- (Accessed July 16, 2025).

NAVCENT Public Affairs. (2024, January 16). *Task Force 59 Launches New Unmanned Task Group 59.1*. Navy. https://www.navy.mil/Press-Office/News-Stories/Article/3645647/task-force-59-launches-new-unmanned-task-group-591/. (Accessed November 5, 2025).

OECD & European Commission (2024). *Health at a Glance: Europe 2024: State of Health in the EU Cycle*. OECD Publishing. Paris.

OECD. (2024*a*). *AI In Health Huge Potential, Huge Risks*. OECD. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/01/ai-in-health-huge-potential-huge-risks_ff823a24/2f709270-en.pdf.

OECD. (2024*b*). *AI, Data Governance And Privacy Synergies And Areas Of International Co-Operation*. OECD Artificial Intelligence Papers, No. 22, OECD Publishing, Paris.

OECD. (n.d.). *Robustness, security and safety (Principle 1.4).* OECD.AI. https://oecd.ai/en/dashboards/ai-principles/P8. (Accessed January 31, 2026).

OECD. (n.d.*b*). *AI principles*. OECD. https://www.oecd.org/en/topics/ai-principles.html. (Accessed February 4, 2026).

Perrigo, B. (2024, May 21). *No One Truly Knows How AI Systems Work. A New Discovery Could Change That.* Time. https://time.com/6980210/anthropic-interpretability-ai-safety-research/ (Accessed 25 June 2025).

Pettyjohn, S. (2024, February 8). *Evolution Not Revolution Drone Warfare in Russia's 2022 Invasion of Ukraine*. Center for a New American Security. https://www.cnas.org/publications/reports/evolution-not-revolution.

Powell, R. (2024, July 31). *The EU AI Act: National Security Implications*. Centre for Emerging Technologies and Security. https://cetas.turing.ac.uk/publications/eu-ai-act-national-security-implications. (Accessed January 29, 2026).

Qureshi, R., Irfan, M., Gondal, T. M., Khan, S., Wu, J., Hadi, M. U., Heymach, J., Le, X., Yan, H., & Alam, T. (2023). *AI in drug discovery and its clinical relevance.* Heliyon, *9*(7).

Rai, D. H. (2024). *Artificial Intelligence Through Time: A Comprehensive Historical Review*. Tribhuvan University

Rashid, A. B., Kausik, A. K., Al Hassan Sunny, A., & Bappy, M. H. (2023). *Artificial intelligence in the military: An overview of the capabilities, applications, and challenges.* International journal of intelligent systems.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), Official Journal L 2024/1689.

Renshaw, J., & Hunnicutt, T. (2024, November 17). *Biden, Xi agree that humans, not AI, should control nuclear arms.* Reuters. https://www.reuters.com/world/biden-xi-agreed-that-humans-not-ai-should-control-nuclear-weapons-white-house-2024-11-16/. (Accessed January 29, 2026).

Rickli, J., & Mantellassi, F. (2023). *Artificial intelligence in warfare. Military Uses of AI and Their International Security Implications.* In Raska, M., & Bitzinger, R. A. (Eds.). *The AI wave in defence innovation: Assessing military artificial intelligence strategies, capabilities, and trajectories*. Taylor & Francis.

Roulette, J., Bryan-Low, C., & Balmforth, T., (2025, July 25*). Musk ordered shutdown of Starlink satellite service as Ukraine retook territory from Russia*. Reuters. https://www.reuters.com/investigations/musk-ordered-shutdown-starlink-satellite-service-ukraine-retook-territory-russia-2025-07-25/. (Accessed February 5, 2026).

Roumate, F. (2024). *Artificial intelligence and the new world order*. Springer Nature Switzerland.

Santora, M., Jakes, L., Kramer, A. E., Hernandez, M., & Sholudko, L. (2025, March 3). *A Thousand Snipers In The Sky: The New War In Ukraine.* The New York Times. https://www.nytimes.com/interactive/2025/03/03/world/europe/ukraine-russia-war-drones-deaths.html (Accessed July 4, 2025).

Sayler, K. M. (2020). *Artificial Intelligence and National Security*. Congressional Research Service. R45178.

Sharma, S., Rawal, R., & Shah, D. (2023). *Addressing the challenges of AI-based telemedicine: Best practices and lessons learned.* Journal of education and health promotion, 12, 338.

Sheikh, H., Prins, C., & Schrijvers, E. (2023). *Mission AI: The new system technology*. Springer.

Shin, K. Y., Lee, J. K., Kang, K. H., Hong, W. G., & Han, C. H. (2019). *The current applications and future directions of artificial intelligence for military logistics*. Journal of Digital Contents Society, *20*(12), 2433-2444.

Slawomirski, L., Kelly, D., de Bienassis, K., Kallas, K. A., & Klazinga, N. (2025). *The economics of diagnostic safety (No. 176).* OECD Publishing.

Stepanenko, K. (2025, June 2). *The Battlefield AI Revolution Is Not Here Yet: The Status of Current Russian and Ukrainian AI Drone Efforts*. Institute For The Study Of War. https://understandingwar.org/backgrounder/battlefield-ai-revolution-not-here-yet-status-current-russian-and-ukrainian-ai-drone (Accessed July 7, 2025).

Svenmarck, P., Luotsinen, L., Nilsson, M., Schubert, J. (2018). *Possibilities and Challenges for Artificial Intelligence in Military Applications.* In Proceedings of the NATO Big Data and Artificial Intelligence for Military Decision Making Specialists' Meeting. Neuilly-sur-Seine. NATO Research and Technology Organisation. https://www.researchgate.net/publication/326774966_Possibilities_and_Challenges_for_Artificial_Intelligence_in_Military_Applications (Accessed July 5, 2025).

Swetha, T., Kumaran, U., Meena, V. P., & Hameed, I. A. (2025). *Leveraging AI for enhanced cybersecurity: a comprehensive review.* Discover Applied Sciences, 7(6), 584.

Systematic. (2023). *Systematic Lands a Multi-million Euro Contract with the British Army*. https://systematic.com/int/newsroom/corporate-news/systematic-lands-a-multi-million-euro-contract-with-the-british-army/. (Accessed October 18, 2025).

Taddeo, M. (2024). *The ethics of artificial intelligence in defence*. Oxford University Press.

The Economist (2017, May 6). *The World's Most Valuable Resource Is No Longer Oil, but Data*. The Economist. https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data (Accessed 22 June 2025).

The Economist. (2024, July 31). *How America Built an AI Tool To Predict Taliban Attacks*. The Economist. https://www.economist.com/science-and-technology/2024/07/31/how-america-built-an-ai-tool-to-predict-taliban-attacks. (Accessed January 24, 2026).

Tierney, A. A., Gayre, G., Hoberman, B., Mattern, B., Ballesca, M., Kipnis, P., Liu, V. & Lee, K. (2024). *Ambient artificial intelligence scribes to alleviate the burden of clinical documentation*. NEJM Catalyst Innovations in Care Delivery, 5(3), CAT-23.

Turner Lee, N., & West, D.M. (2025, November 5). *The future of data centers*. Brookings. https://www.brookings.edu/articles/the-future-of-data-centers/. (Accessed January 27, 2026).

U. S. Code. (2021). *15 U.S. Code § 9401 – Definitions*. U.S. Code.

U.S. Department of Commerce, Bureau of Industry and Security (2022, October 13). *Implementation of Additional Export Controls: Certain Advanced Computing and Semiconductor Manufacturing Items; Supercomputer and Semiconductor End Use; Entity List Modification*. Federal Register. 87 (197). Fed. Reg. 62186. https://www.govinfo.gov/content/pkg/FR-2022-10-13/pdf/2022-21658.pdf.

U.S. Department of Defense. (2018). *Summary of the 2018 Department of Defense artificial intelligence strategy: Harnessing AI to advance our security and prosperity*. https://media.defense.gov/2019/Feb/12/2002088963/-1/-1/1/SUMMARY-OF-DOD-AI-STRATEGY.PDF.

U.S. Department of Defense. (2022). *Summary of the Joint All-Domain Command and Control (JADC2) strategy*. https://media.defense.gov/2022/Mar/17/2002958406/-1/-1/1/SUMMARY-OF-THE-JOINT-ALL-DOMAIN-COMMAND-AND-CONTROL-STRATEGY.pdf.

U.S. Geological Survey. Mineral Resources Program. (2025). *Key Minerals in Data Centers Infographic*. U.S. Geological Survey. https://www.usgs.gov/media/images/key-minerals-data-centers-infographic. (Accessed January 27, 2026).

United Nations. (n.d.). *Artificial Intelligence (AI).* United Nations. https://www.un.org/en/global-issues/artificial-intelligence (Accessed 21 June 2025).

United Nations Conference on Trade and Development (UNCTAD). (2025). *Technology and Innovation Report Inclusive Artificial Intelligence for Development.* United Nations. https://unctad.org/system/files/official-document/tir2025_en.pdf.

United Nations General Assembly (UNGA). Resolution 79/62 [Lethal autonomous weapons systems]. https://digitallibrary.un.org/record/4065061?v=pdf.

Väänänen, A., Haataja, K., Vehviläinen-Julkunen, K., & Toivanen, P. (2021). *AI in healthcare: A narrative review.* F1000Research, 10, 6.

Van Leeuwen, K. G., de Rooij, M., Schalekamp, S., Van Ginneken, B., & Rutten, M. J. (2022). *How does artificial intelligence in radiology improve efficiency and health outcomes?.* Pediatric radiology, 52(11).

Verbruggen, M. (2022). *No, Not That Veri cation: Challenges Posed by Testing, Evaluation, Validation and Verification of Artificial Intelligence in Weapon*. In Reinhold, T. & Schörnig, N. (Eds.). *Armament, Arms Control and Artificial Intelligence. The Janus-faced Nature of Machine Learning in the Military Realm System*. Springer.

Walter, Y. (2024). *Managing the race to the moon: Global policy and governance in Artificial Intelligence regulation—A contemporary overview and an analysis of socioeconomic consequences.* Discover Artificial Intelligence, 4(1), 14.

Wasserman, L., & Wasserman, Y. (2022). *Hospital cybersecurity risks and gaps: Review (for the non-cyber professional)*. Frontiers in digital health, 4, 862221.

Wearden, G., & Stewart, H. (2026, January 23). Young will suffer most when AI 'tsunami' hits jobs, says head of IMF. The Guardian. https://www.theguardian.com/technology/2026/jan/23/ai-tsunami-labour-market-youth-employment-says-head-of-imf-davos (Accessed January 24, 2026).

Whyte, C. A., Thrall, T., & Mazanec, B. M., (2020). *IWIO in the Age of Cyber Conflict*. Routledge. 344. Oxfordshire, UK.

Williams, G. S., Koua, E. L., Abdelmalik, P., Kambale, F., Kibangou, E., Nguna, J., Okot, C., Akpan, G., Moussana, F., Kimenyi, J. P., Zaza, R., Carerra, R. M., Rabiyan, Y., Woolhouse, M., Okeibunor, J., & Gueye, A. S. (2025). *Evaluation of the epidemic intelligence from open sources (EIOS) system for the early detection of outbreaks and health emergencies in the African region.* BMC public health. 25(1), 857.

World Health Organization. (2016). *International Health Regulations (2005) (3rd ed.).* WHO.

World Health Organization. (2021). *Ethics and governance of artificial intelligence for health: WHO guidance.* WHO. Geneva.

World Health Organization. (n.d.). *Health Security*. WHO. https://www.who.int/health-topics/health-security#tab=tab_1 (Accessed January 6, 2026).

World Health Organization. (n.d.*b*). *Artificial intelligence for health emergencies: WHO advances public health intelligence and surveillance through innovation*. WHO. https://www.emro.who.int/media/news/artificial-intelligence-for-health-emergencies-who-advances-public-health-intelligence-and-surveillance-through-innovation.html. (Accessed February 5, 2026).

Wright, S. (2006). *Terrorists and biological weapons. Forging the linkage in the Clinton Administration*. Politics and the life sciences: the journal of the Association for Politics and the Life Sciences, 25(1-2), 57–115.

Zhang, H., Khanal,S., & Taeihagh, A. (2025) *Public-Private Power plays in Generative AI Era: Balancing Big Tech Regulation Amidst Global AI Race*. Digital Government: Research and Practice. Vol 6, 2, Article 26

Zou, M., & Zhang, L. (2025). *Navigating China's regulatory approach to generative artificial intelligence and large language models*. Cambridge Forum on AI: Law and Governance, 1, e8. doi:10.1017/cfl.2024.4

Zysk, K. (2024). *High Hopes Amid Hard Realities: Defense AI in Russia.* In Borchert, H., Schütz, T., & Verbovszky, J. (Eds). (2024). *The Very Long Game. 25 Case Studies on the Global State of Defense AI.* Springer.