Prajesh Mishra

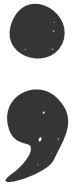
Strategic Intelligence

Artificial Intelligence, Cyber Defense, and Security in the Digital Age

DeepScience

Strategic Intelligence: Artificial Intelligence, Cyber Defense, and Security in the Digital Age

Prajesh Mishra



DeepScience

Published, marketed, and distributed by:

Deep Science Publishing, 2025
USA | UK | India | Turkey
Reg. No. MH-33-0523625
www.deepscienceresearch.com
editor@deepscienceresearch.com
WhatsApp: +91 7977171947

ISBN: 978-93-7185-094-0

E-ISBN: 978-93-7185-637-9

https://doi.org/10.70593/978-93-7185-637-9

Citation: Mishra, P. (2025). *Strategic Intelligence: Artificial Intelligence, Cyber Defense, and Security in the Digital Age*. Deep Science Publishing. https://doi.org/10.70593/978-93-7185-637-9

Preface

The world has never witnessed the basis of national security to be redefined in ways that Artificial Intelligence is doing. Autonomous systems, real-time analytics and scalable cyber protection technologies used to take milliseconds to identify potential attacks that otherwise took full teams of human analysts to identify and process before confirming their presence and need intercession. To the United States, this change is a turning point as not only a turning point in technology, but also in strategy.

The book discusses the way in which AI is redefining contemporary conflict in the cyberspace, intelligence, infrastructure and geopolitics. It delves into the potential and the threat: self-policing AI able to pick up threats quicker than any mortal and antagonistic AI able to use the system vulnerabilities, faster and more extensively than ever before. The prospects of national defense are changing at an extremely fast pace along with deepfakes and misinformation to digital-twin cybersecurity and autonomous battlefield systems.

Based on studies and experience related to AI, cybersecurity, and real-time data analytics, the book seeks to offer a simple and easy to use conceptualization on how these changes would be comprehended. It is addressed to the technologists, policymakers, students, and readers who may be keen on the role of AI in the future of American security.

We are starting a smart period where national power can be as reliant on digital fortitude as it can be on physical prowess. Our attitude towards the moral and artificial intelligence, our justice and power, our innovativeness and collaboration are the choices and actions we make today which will determine whether AI will serve as our greatest strength or largest weakness.

I hope that this piece of work can shed light on the way ahead.

Prajesh Mishra

Table of Contents

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 1: The New Era of Intelligent Conflict: AI as Strategic Sovereignty

1. Introduction

Since the dawn of nuclear weapons, war has been conceived and carried out as a contest of wills. While retaining this element throughout the historical evolution toward the current era of relatively rare but increasingly threatening large-scale wars between great powers, the warfighting focus has shifted decisively away from the mere clash of armed forces toward the careful calculation of strategic advantage. The development of the world-leading ability of the United States to leverage advanced technology and intelligence to eliminate high-value, low-risk targets with precision long range strikes reduced the primary challenge to its prevailing deterrent posture from coping with its rivals' attempts to prepare lists of bombs for missiles to support Armageddon strikes to the still-formidable task of ensuring that its enemies lacked the capability to eventually send the planned strike into the air. Algorithmic warfare represents a paradigm shift in this understanding of war as a conflict of wills. War is not only a foreboding possibility that no one wants to initiate, but is now front of mind for decision-makers in AI-fortified defense establishments seeking to position themselves so that it can be won with the lowest possible cost in time, money, and lives.

This has resulted in a shift in focus away from the weapons and their synergistic combinations toward the hearts and minds of those planning and executing the strikes, steering more effort toward making better decisions faster than the adversary than toward making the best decision. Recognizing for the first time that timing is the ultimate arbiter of operational success and failure, algorithmic warfare is bringing disparate yet parallel areas of national security—the artificial intelligence revolution in computer science, citizen safety, classified information sharing, active cyber defense and liberal democratic resilience—closer together. Driven by the rapid rise of intelligent conflict, these national security domains are experiencing twin waves of change. The first involves making them fit for purpose long before the onset of war to endure and prevail in the face of new and commensurately rising threats. The second centres on speeding up the proverbial military expeditions of the United States to ensure it retains the edge not only in major wars but also all forms of military conflict.

2. From Traditional War to Algorithmic Warfare

The phrase paradigm shift refers to a drastic shift in the way of mental models, both in the scale and duration of mental models like the transition between the Ptolemaic and Copernican worldview or the Industrial Revolution. In case of war, the term describes fewer and more tangible changes in the way they are waged.. Such a paradigmatic war proposal is not straining under the term when contrasted with earlier warfare capabilities and methods. Traditional warfare involves roughly symmetrical force-on-force engagements, whereas algorithmic warfare features autonomously and massively scalable algorithmic systems optimized for edge capability and speed and employing gaming theory to streamline project,

function, and behavior. Algorithmic warfare functionality has matured over the past decade, and even though it remains an incomplete paradigmatic war, multiple research, development, and operational vehicles – as well as incidents hinting at its power – abound [1].

The creation of the comprehensive map of operations reveals the major changes in the operations sphere, whereby the evidence of various sources should be provided to map contest space. Cases in point indicate that margin control and subtle offense and defense can be making success on redefining success in algorithmic contests and have far reaching consequences in deterrence and arms control. This concept of CySuite as a competitive strategic cycle of interdependence over multiple levels of success and failure of campaigns in the security domain, but still does not take into account the particularities of Algorithmic War CySuite where diversion logically decays, misinformation flourishes, and denial should be stretched through the entire life cycle in the project in case full project denial is desirable.

Ethical and legal accountability considerations are leaping to the forefront of the discourse in Algorithmic Warfare, where evidence of a Human-in-the-Loop responsibility paradigm is lacking. International warfare law provides a guiding framework, and the principles of proportionality, distinction, and necessity remain valid. The exponentially increasing argument for Algorithmic Warfare CySuite, however, opens a slippery slope that bears further risk, and even though this fallacy may seem exaggerated, the intuition remains practically valid.

2.1. Conceptual shift in warfare paradigms

As an illustration of the burgeoning consequences of the technology-driving Fourth Industrial Revolution, the operational relevance and effectiveness of combat action have burgeoned during recent decades and more recently ebbed with respect to war-fighting and contestation between states and major powers. These phenomena are now part of the private sector-dominated and AI-powered Fourth New Industrial Revolution Era. In a 21st-century context, such socio-political and technological conditions and trends denote a shift toward Algorithmic Warfare, the disruptive alternative of intelligent conflict, and a consequential new operational approach termed Intelligent War. Such a paradigm shift represents a change in the actual conditions, environment, and conduct, rather than merely the tools and means. Indeed, Algorithmic Warfare forms an integral aspect of Intelligent Conflict, typifying one of the three dimensions of such intelligent conflict. A broad range of commercial actors are developing, augmenting, or enabling emergent military technologies. Dual-use capabilities are proliferating. Consequently, the private sector and outside stakeholders are now often the principal or indeed only actors in many combat missions, operations, and domains.

Vehicle Technologies and Operations offer a case study of these transformations. A product of the technology-driving Fourth Industrial Revolution, the operational relevance and effectiveness of these capabilities have burgeoned during recent decades and more recently ebbed with respect to war-fighting and contestation between states and major powers [1-3]. Air vehicles and operations epitomised a prior transformation in operational capabilities. Such implementation has now accelerated through Algorithmic Warfare and Intelligent Conflict. The basic transformation is reflected in development, enhancement as well as utilization of operational capabilities by the private sector. CIA Industry 4.0, Data-Processing-Structured, and civilian (commercial) UAV and AAM technologies and operations have turned out to be central to land, surface, air, and maritime operations.

2.2. Operational transformations and case examples

Across societies that field militaries, large-scale armed conflict is entering an operational phase with arbitrary boundaries, ill-defined direction of travel, and unparalleled volumes of death and destructionIt extends previous and smaller forms of manifestation in the Syrian Civil War, the Second Nagorno-Karabakh War, and the Russo-Ukrainian War; it can simply be defined as automated and automated combat, definitory historical description of the phase is a paradigm shift towards algorithmic warfare. Major events across a spectrum of conflicts involving a rich dose of tactical and operational sophistication—such as the extensive use of drones by both sides in the ongoing war between Russia and Ukraine; the re-politicization of korsar unmanned surface vessels, including the first-ever strike on the head office of Russia's Black Sea Fleet with surface drones; and imitation of sophisticated algorithms as espoused by Hezbollah—probably show that the shiny objects of automation and autonomy are merely tools to current and future conflict.

The thin edges of campaigns of the past, which involved operation and war actions, seem to subside in a certain manner, which is marked by the disastrous losses that Russia experienced following years of planning the invasion of Ukraine. but events in Ukraine are not merely demonstrating the blastococcal effects of algorithmically facilitated-high-intensity asymmetrical war fighting, but they are also demonstrating its permanence and permanence limits of ethicality as well. The extravagant wealth in degree-zero brutalization of the civil-armed grey zone and consequently the ethical genres of retribution is now being played out on a global stage; individuals and organizations spanning the panorama of race and belief systems are joining, directly or indirectly, the one-sided sanctioning effort against Russia that desires to dissolve the sense of diplomatic propriety and set a precedent for future civil conflict resolution.

2.3. Ethical and legal considerations in algorithmic combat

A portion of the 2024 Guide to AI and National Security, which is prepared by the Office of the Secretary of Defense, is on the implications of AI on operable conduct. Though the field of armed conflict has traditionally been an area to which the international humanitarian law applies, the increasing automation of warfare poses a myriad of ethical, legal, and operational issues concerning the ways of detection, targeting, and engagement, though the underlying Ash Panette monograph fields areas of ethical hazard, both substantively and procedurally.

Even the guide itself says that facilitating the occurrence of algorithmic combat in line with the laws of war and the Department itself are paramount, but does not stipulate that complete control of sensor and shooter remains in human hands. Notably, the DOD is yet to promulgate particular criteria of harm algorithmically inflicted. They may encompass customary principles, such as distinction, proportionality, necessity, and accountability or integrate Pillar 1 of The U.N. Secretary-General road map to digital cooperation with the Guiding Principles on Business and Human Rights as the basis of evidence-based design decisions. These standards make algorithmic combat an implementation of the laws of war with the help of an algorithm in the decision-making process and the decision to act..

3. AI as the New Strategic Asset

Like nuclear materials or cryptographic superiority, military prowess in the modern era is defined by access to the best and most AI systems. These capabilities are at once technical and operational: performance in AI must be considered in conjunction with a host of national resources – from processing, data, and talent sufficiency to problem definition, access to real-world distribution, and, in time, problem

domain itself. The quality of these factors ultimately determines the utility of national AI supremacy as a strategic resource. Given the increasing volatility and reduction of margins in algorithmic warfare, however, measures of performance in AI are also increasingly capable of direct, signal-dependent interpretation.

In addition to the suppliers of foundational and operational technology, the geographical centers of critical advances in AI are also rapidly evolving into matters of strategic concern. On the one hand, security of supply chains of capital-intensive parts and particularly semiconductors and graphic processing units is of the essence in the short-term with any dual-purpose technology. Out of the short-term perspective, the accessibility of supportive settings to the building of advanced AI systems, particularly appealing to highly mobile global labour force, becomes determinant. In a longer time horizon that moves beneath the 24/7 nature of the TiVo nation, investment in strategy-compliant AI-related assets becomes key.

3.1. Comparative appraisal with historical strategic resources

In the Fourth Industrial Revolution, characterized by the emergence of artificial intelligence (AI), global politics is undergoing changes that are increasingly akin to established logic on resources of strategic importance: factors Fischer proposes for which there is relative scarcity, for which there is basic need, and assets whose importance and value is growing. AI resources manifest characteristics comparable to a form of oil in a digital battlefield where the economy operates at the nanosecond and at the same time a form of uranium associated with the nuclear weapons of the twenty-first century. While the term revolution even appears in association with AI digitization, the emphasis now focuses less on characterization and more on the direction of evolution. Resources, indicators of superiority, risk represented by asymmetric dependencies, supply chain configuration and demand,able factors of production of the Axel Springer Gmbh-Gruppe and the availability of excellence suppliers, area by area, are, day by day, increasingly of greater interest. In an era in which visibility, and survival are linked respectively to a higher speed of processing and the capacity to react in significantly shorter times than those for an apparently well-planned decision, the extreme speed of data processing - in many applications far exceeding that of the human brain - and the access to an exponentially higher quantity have ao become resources.

Following these reasoning lines emphasizes that strategic advantages will be based on specific indicators areas in which superpower AI resources are strengthen. The sources of information for analysis are therefore partly different, with particular emphasis on the aspect of data exploited by the digital economy. The ability of advanced mathematics - both by experts and by software - to select and weigh information source by source for quality, credence and utility greatly increases the overall potential of intelligence, Beyond necessarily becoming highly dominant - often a risk given the number of smart concepts developed even in a short time - it is the broad based and coherence-linking multiplicity of AI analysis that proves decisive. Even defence decision factors drone rely to AI sophisticated risk-checking to decide, with consequent delocalizations- giving up automatic decision-making or even delegation to proposals without external supervision.

3.2. Indicators of AI superiority and national capability

The strategic consequences of AI arms race and its scarcity, economic importance, and urgency substantiate the difference of an era when the ability of a country to turn the AI benefits into the relevant

military solutions becomes dominant. Competition based on AI weapons creates new war precursors capable of eliminating the safety margins that existed in historical times due to second-strike capabilities and mutual-assured destruction capabilities. Furthermore, containment that merely emphasizes on denial will also allow certain states to have competitive advantages in the AI-related fields other than the national defense. The broad-based way of AI strategic evaluation, therefore, must involve tracking the happening of parameters of performance that incorporates possessing the immediate consequence of prioritized investments to AI growth in addition to the logistical framework which converts the investments to victories applied.

Of key interest is the imbalanced nature of the most competent nation on the one hand and the most competent competitor on the other hand and the distribution of winning in different countries. There are particular signs of relative performance these include the presence of the four major engines of AI development, availability, and production which include the enhanced processing power, the presence of large and dependable data sets, capable and rich talent pool, and the large-scale development of AI applications in terms of quantity, quality, areas of application and geographical coverage. The fact that any one of the states, including allies of the U.S., who hold an indirect containment role, is able to continue to have relative access to any single engine, especially the major nodes in supply-chains, that support large scale processing power and contain available talent pools, further complicates the analysis.

3.3. Supply chains, talent, and infrastructure for AI dominance

The capability to scale to the level of surviving and operating on military-intelligent algorithms is dependent on three interdependent national qualities: a robust supply chain of AI-enabling hardware; a viable, sufficient, and representative amount of data, including training data, reinforcement learning ground truth, and synthetic simulation modeled reality; and reliable infrastructure that will support a sequence of failed experiments over months and years.

The scale of semiconductor manufacture, especially of graphics-processing units (GPUs) and field programmable gate arrays (FPGAs), is the most vital factor of supply chain, and the trend is in favor of the ones with the high performance to price ratio. As much cutting-edge technology as possible must remain available to military- and intelligence-community customers, including system-on-chip (SoC) packages [2,4]. Cyber-physical dependencies are of immense concern, giving extra importance to satisfying civilian demand through push-and-pull policies that minimize market friction. Stimulating demand is key, especially for calibration and other low-level processes where apprenticeships are possible, helping to build out the domestic talent ecosystem.

Talented AI developers are scarce and prized. In the short term, quantitative factors such as the size and quality of the programming workforce combined with a friendly immigration policy support AI capability more than qualitative considerations such as funding or brain drain. However, as the opportunity for public- and special-sector employment submerges matchless-individual strategic advantages, the long-term balance of the contending training, recruiting, and retention systems plays a more decisive role.

AI training, development, and testing infrastructures for military-intelligence purposes are clearer. Generating Artificial Intelligence in-the-loop is the primary concern: combat-tested battle data is the ultimate scarce resource, and gaming simulation environments are imperfect. Training AI in-the-loop during conventional, adaptive, shaping, and counter-insurgency parts of conflict is a means to an end—the Armageddon with scarce-transitory advantages in a potential opponent’s strategic center of gravity (be

it decision-making, military-industrial supply chains, or public will). Cyber-physical training-enabling infrastructure within the Five Eyes community must be sited to resist-persist against whichever pair of players' dry-ice-deployed Advanced Cyber-Physical Patrolds first strike first.

4. AI, Intelligence, and Defense Innovation

The full-spectrum artificial intelligence carries out intelligence collection and operational evaluation to assemble a host of data to build credible and timely situational analysis to commanders. Artificial intelligence systems are becoming real time situationally aware which means that they can make quick and independent decisions. In cyberspace and space domain, offensive operations involving system development, production, and deployment in scale are being performed based on automation and AI. Besides, AI is being exploited to improve the capabilities of cross-domain operations enabling the deployment of land, air, sea, cyber and space where the commanders have infinite options to use with regards to their operations. The pace of development of algorithms is superior to the traditional defense innovation process. To redesign their defense sectors, structures of governance, and procurements, governments are seeking to follow AI-first solutions.

Historically, pan-domain intelligence gathering with direct support to operations has been the task of Intelligence Agencies, the operational analysis of intelligence inputs has similarly resided within the domain of Agency-led Capability Development Groups, and the final operational assessment with respect to mission objectives within the Command Structure. However, the speed of algorithm development is outpacing the traditional defence innovation cycle, presenting a variety of challenges, risks and new opportunities that traditional command structures struggle to assess in a timely manner. AI-enabled support to intelligence gathering and analysis provides an ideal foundation for full-spectrum support to defence operations, able to exploit and fuse a multitude of open source data streams to provide commanders with rich, credible and timely operational assessments.

4.1. AI-enabled intelligence gathering and analysis

Advanced AI techniques facilitate rapid, efficient, and comprehensive information gathering and analysis. AI can comb through open-source information, proprietary intelligence, and other information to greatly accelerate the intelligence cycle and enhance the quality of intelligence products with novel insights. AI can automate collecting information from diverse and incomplete sources, including existing SIGINT systems, to help detect and analyze signs of diversion, including cyber support to offensive use, rapid military preparations, and preparations for disinformation campaigns. Multi-fidelity/ distributed corroboration of open sources (including human sources) can help in establishing credibility and building confidence.

Beyond dramatically accelerating the identification of candidates at risk of engaging in forbidden behavior in general with high precision, these technologies may help support fusion of SIGINT and associated geographical intelligence with human-source intelligence on acts of violence, thus addressing some of the credibility assessment problems traditional analysts face when evaluating contradictory-SIGINT and human-sources reporting. Logic-based reasoning of mixed degree of truth level with probabilistic support can provide a high-performance analysis toolset for rapidly pinpointing the locations of nuclear warheads, ballistic missiles, and related ammunition during peacetime and for capitalizing on such information to establish the credibility of human-sourced reporting when key elements are inconsistent with sampling cross-recovery of SIGINT [5,6].

4.2. Autonomy, decision-making, and risk

The main factor to consider in the impact of AI on the autonomous nature of military operations is the way the resulting systems will decide how the operations are run. Even the strategic operating environment in the future is likely to have AI combat systems with different degrees of autonomy.. Some may operate in a fully autonomous mode, that is, without direct human control, while others will include a human in the loop who is directly involved in certain stages. Others may be fully autonomous in operation i.e. they do not require a human intervention whereas others will involve a human in the loop who is directly involved in some parts[6-7]. Other individuals might submit intended actions to be approved or may offer their forecasted plan of action during decision-support mode, so that a human operator can be able to give more judgement on the situation and often give guidance. Both the levels of autonomy possess strengths and weaknesses that might vary with regard to the task and operating background.

In the decision-making of autonomous execution, when these decisions concern sending a missile attack at the target, assembling information on a new Internet threat organization or collecting information on a large-scale business port, the entire responsibility moves off the shoulders of humanity, and threats of military accident are redirected onto other people.

4.3. Cyber, space, and cross-domain implications

The development of AI will have a high probability of influencing space conflict in the cyber space, outer space, and other spheres. Risks of an error in calculation and unwanted escalation in case of deterrence failure may arise as a result of more automation and the growing complexity of the contemporary conflict. Each party will desire to convince the opposing parties that it still has deterrence and at the same time apply gestures of competence and determination to influence what the adversaries think. Signaling and assurance strategies will have to consider the speed of operation, and failure to identify, nurture, and evaluate credible signals.

When developing new forms of escalation and their consequences to the strategic stability, the appearance of the new forms should also be understood with serious consideration. Not only will new AI capabilities make existing targets more vulnerable, they may also create new ones, leading to the possibility of an adversary attempting to retaliate in nontraditional domains. In addition to this, the growing capacity of producing and feeding misleading-appearing-information and media content, exposes the threat of generating large-scale and speedy-created information warfare of deception and deception-supporting information, which have the potential to make the adversary militarily off track with the perception of information warfare on a grand scale[6-9]. The combination of the increasing scale, speed, and complexity of operations creates potential implications for Gray-Taylor dynamics that require serious thought, especially when combined with increasing levels of autonomy in military operations. The increasing numbers of automated systems will also likely raise the risks of actual use through a natural tendency toward automation bias, even in the absence of deliberate escalation.

5. Strategic Stability and AI Arms Considerations

In the era of very high rates of automation deterrence must be based on crisis signalling and promise at an environment of fast-growing capabilities, realistic evaluation of the dangers of escalations, and believable second-strike capabilities of an increasing number of nuclear countries. Assumptions of deterrence in predictability, decision time, risk tolerance can need to be overhauled. The AI-enabled systems can be

able to improve the stability of the crisis by providing more reliable, quicker, and less expensive information sharing. Nevertheless, the rapid development of these technologies also leads to generous stocks of low-cost resources and increases the risks associated with assurance.

Proposals on norms and arms-control that are associated with crisis stability, offensive-autonomy boundaries, and the risk-control AI representations are urgent. One should work at creation of a low level transparency of decision-assist systems as well as shared operational space and environment so that human-in-the-loop solutions become possible. The strategic partnerships should take the aspect of cross domain interoperability and joint deterrence in the face of unpredictable competitors. It is urgent to find a solution to the problem of AI verification and to rely on one of the partners in an alliance as a source of necessary technologies is perilous.

The emergence of AI as a strategic asset changes the calculation of the needed quantities on most of the conventional deterrent capabilities. The AI enabled deterrence is reliant on speeding up innovation cycles; being more willing to take risk, in case of a crisis or a conflict; being more affordable and possibly less reliable, in terms of escalation controls; and having shorter warning time. The safest methods are supply-chain assurance and the establishment of primary product stocks that deal with the second-order dependencies.

5.1. Deterrence in an age of rapid automation

Rapidity of the ongoing transition to algorithmic warfare may dilute deterrent credibility. Algorithms in conflict do not have intrinsic moral agency and neither are those operating them. Notion of rational actor is often an assumption in game-theory based analyses. Functionally stupidity may yet replicate results of collective stupidity without rational authority. Consequently, deterrence strategy incorporating human managed Algorithm scenario may prove flawed in accuracy. Emphasis in crisis management therefore must shift from deterrence to assurance[10-12]. Yet transformation to algorithmic warfare remains an evolution in character and not nature. Algorithms enhance the efficiency of systems not in their near-term kinetics but in the longer run placements and deployments. In that hinterland both sides still shoal together when Algorithm. Yet here deterrence of the erstwhile nature is vital—to credibly against efforts at co-existence. Plurality of Algorithms-and-a-Command heuristics add dimensionality yet makes coverage of all-dimensionality difficult.

Algorithms determine decisions but rapidly verifying Algorithm tutorials are non-existent. Deterring some residing closer to irrationality may snowball into a catastrophe. Such also is the nature of dispersed data velocity in such milieu with falsehood dressed as truth. Such negative dynamics demand greater attention in dispelling doubts within the rival/competing clusters. During the Cuban Missile Crisis of 1962, U.S. sources were aware the Soviet decision-making was under high sensor coverage/trackability. Yet an assessment made by a child of God/Peacenik—Dwight D. Eisenhower—before his exit believed greater emphasis should be laid on not simply deterring disaster from striking but ensuring confidence that disaster was not going to strike. Rapid algorithmic automateabilité in some sectors can lead to an elongated Assimilation time [5-8]. The resulting slack in pace and scale of weapon systems used especially nuclear known as second-strike or retaliation capability is stark and continues to worsen. Misdirections during such phases of a crisis cycle can rapidly connect the dots into catastrophe.

5.2. Arms control, norms, and governance mechanisms

AI makes arms control more complicated, particularly in the fields where there is accelerating automation. Deterrence generally involves reassuring second strike capabilities, crisis communication and stating intentions. Under such a dynamic scenario, the innovation cycle moves very quickly thus obscuring these aspects of deterrence as the opportunities exploitable by them may not be apparent to the opposition. To keep the power of deterrence functioning throughout the competition range, especially at the time when the dangers of miscalculation are also increasing, the use of AI needs to be offered on a reliable and credible basis. An example of such deterrence with promptly maturing autonomous capabilities might need closer coordination than the one of non-autonomous sets of capabilities.

In addition to deterrence, AI development opens prospects of arms control, governance, as well as norm-building. Governance depends on reviewing of independent decisions including taking acceptable risks in making loss-of-life decisions. Stability might also be supported through the mechanism of enhancing transparency and predictability regarding algorithms of concern. Special check-up procedures of high-impact systems that potentially contain destabilizing aspects may also be helpful in eliminating the risk of unintended escalation or conflict..

5.3. Strategic alliances and AI interoperability

Dynamics of deterrence is changing in a world that is characterized by a high rate of automation of military actions. The classical ideas need to be re-considered especially such significant characteristics of these activities as assurance, signaling and ambiguity. Change in the tempo of military activity can reduce the effectiveness of deterrence by rendering traditional signals of resolve or intent less effective; rapid, unpredictable, and (to a degree) opportunistic operations can blur and obscure the stabilizing properties of major military postures[7,13-16]. These factors cast doubt on the reliability of deterrent signals, complicating efforts to assure partners and reassure adversaries.

Alliances and partnerships are often proposed as means of coping with new forms of deterrence. Such arrangements are prone to reciprocal dependency. Should they succeed in generating truly interoperable capabilities—especially with regard to AI-enabled systems and operations—these systems can become the focus of strategic tensions either by virtue of their empowering influence or by presenting highly tempting targets for a prospective adversary. Achieving a sufficient degree of readiness to capitalize on a new-found capacity for rapid automation, especially to permit greater survivability of offensive-force operations, is also being recognized as a prerequisite for stable partnerships.

6. National Security Policy Implications

Sustained standard-setting requires careful attention to national security policy. In the interests of advanced AI applications, decisionmakers should consider reforming government and industry organizations to improve the governance of AI and related technologies, forming the right teams to harness AI for intelligence operations, refining procurement processes, and investing in talent development. These endeavors must be supported by accepted standards of AI excellence that extend past imaging and processing features, metrics of performance on quality of data, and red-teaming activities aimed at establishing the vulnerabilities of the system. Governmental, industry, and academic collaboration and close affiliation with non-profit organizations could assist in making sure that the AI development process is conducted in an ethical direction and that the technologies are used by the society in a human-friendly way.

Formal national security organizations should not shoulder the responsibility for sanctions and risk assessment alone. Industry, academia, and the broader community also have roles in better understanding security implications. All stakeholders should work together to establish validation and assurance standards that promote secure technologies and instill public confidence, reduce the chances of malicious use and criminal exploitation, and engender a growing awareness of the ethics of machine-facilitated human and social interaction.

6.1. Organizational reform and capability investment

The national leaders have to centralize intelligence, defense development, and Homeland security on AI development. As much as the current systems are already well skilled in their work, they are unable to keep up or outsmart the fast changing threats and opportunities that could occur in AI. New coordination capacity is required to provide actionable information and combined solutions to operators, serve as information to drive acquisition and investment approaches, and inculcate a culture that will allow organizations to complete AI data-hungry, attention-intensive tasks. Differentiated acquisition procedures, (department of defense, intelligence community and department of the homeland security), make the technology insertion and capability development complicated. Such necessary supporting processes can be reinvented and aligned in order to make national security organizations better absorb innovation made by the general U.S. technology ecosystem.

Scaling and increasing the pace of implementing AI in operations would need data in large quantities, which reflects the working conditions and the abilities and strategies of the opponent. The quality of the data should be benchmarked by organizations in the defense and intelligence and homeland security communities to train AI. The quality and scale of data should be considered good and they should be made available to the wider community of developers [2,17-19]. These organizations are also encouraged to develop frequent, realistic red-team training, which are threat emulations intended to simulate the operations and systems that a certain enterprise can expect to face in a given time period, specifically to provide adversarial data of quality to AI models. Having a clearly defined adversary provides a better reference point upon which to assess training data and helps to expose gaps in a given format prior to operational deployment.

6.2. Metrics of AI superiority and assessment frameworks

Measurement of AI capability development is essential for the United States to maintain an advantage in an area recognized as critical for national security. Although over a dozen reports have called for an enhanced focus on examining the character and relevance of a state's AI capacity, no consensus has emerged. The indicators used to evaluate AI development differ across these publications, partly due to their different areas of focus. Requirements for intelligence systems differ from those for weapons systems. AI systems that have evolving commercial deployment are different geometrically and geographically as compared to military development. The ability of the state to transform into AI, integrate it, and implement it in practice will have a significant difference with the scale of its technical talent source. And yet, despite the underlying variations in AI, all military and political leaders recognize the need to study the driving determinants of the resources powering the most capable AI clusters.

It is suggested to measure advanced processing capacities, absorbance of data provided in natural language, ready-to-use datasets, provided with confirmed quality labels, and talent-stock and -flow metrics of AI. There is a national performing deployment and application proposed hardware and

software infrastructures to pay certain attention to. Being represented by the indicators in the form of appropriate absolute or relative measurements, they constitute an approach of measure-of-measurements to the measurement of AI capabilities of states and accordingly competitive balance- a point of interest that provides an understanding of the weaknesses in investment strategies that may lead to a gap in the achievement of AI capacity.. To this end, measuring and using magnitude and quality correlations of commonly employed AI and resource traffic indicators with the realized balance of applied AI capability offer qualifiable insight into national capability. AI-supporting supply chains that are similarly labelled collate and examine risk mitigation, resilience, return on investment, and related measures.

6.3. Ethical, legal, and human-centered design imperatives

Ethics, legality, and the human experience of using, being affected by, and designing AI systems in Defence and National Security therefore cannot be left to ethics committees or the legal system to assess service-by-service, application-by-application. Since military operations are meant to support civilian life, the expectation for human life, agency and oversight in the use, especially, of AI is a matter of human-centered design [3,20]. The purpose of the design in this respect is high usability levels to the targeted user, high levels of transparency to the individuals and communities that are involved, and provision of tools that will serve the purpose of the mission in leaving the broadest range of human choice and agency through the whole process of development and the operational life.

To achieve this vision, every project and every system and stored dataset needs the endorsement of the intended users, people likely to be affected, and developers with complementary expertise and experience. Inclusive design that gives priority to the advice of affected persons is an important step in the right direction; human-centred design implemented at scale for both acquisition and use is the goal. The focus is not only on the design of complex systems but also, importantly, of all the building blocks that go into those systems, that wider view of design leads to the best choices in the use of non-AI technology for Defence and National Security, especially in future-proof and adaptable support of assisted-human operations.

7. Conclusion

To avoid misunderstanding, it should be emphasized that AI does not in itself change the fundamental strategic issues of war. However, given its strategic salience, other actors with the resources and ambition can be expected to devote significant attention to its development. The development, design, and implementation of AI technologies that can coordinate the fighting by making decisions based on algorithms at a scale of time and precision that is millions of times higher than that of humans represent a new trend in international security and strategic stability of every state, and the New War is the war in the era of a fast AI every state, region, and alliance of which has a geographical, regional, and alliance-transcending implication. When AI is applied at scale to military intelligence processes and for key defense innovation tasks, to sectors such as cyber and outer space, and through combat assets, both offense and defense can be made faster, delayered, more precise, and more successful than at any previous time in international history. Although these changes may decrease the threshold of war winning, their implications about deterrence and the worsening of the conflict are less clear and obvious.

An AI arms race is neither a good thing nor can and must be regulated yet, but nonetheless, states continue to devise norms, armament controls and governmental procedures that will reduce the risks associated with automation and AI as well as with algorithmic warfar. Before those measures can be

developed, however, nations must first identify their operational and strategic significance in order to better understand how to control them. Such a nuanced understanding of strategic stability must also recognize that continuous and dynamic competition frequently poses a greater danger than war itself. During periods of rapid innovation and operational future-orientated change, there may be increased temptation for either party to realize an underlying strategic advantage through the threat or actual use of force.

References

[1] Ali O, Abdelbaki W, Shrestha A, Elbasi E, Alryalat MA, Dwivedi YK. A systematic literature review of artificial intelligence in the healthcare sector: Benefits, challenges, methodologies, and functionalities. Journal of Innovation & Knowledge. 2023 Jan 1;8(1):100333.

[2] Zhang K, Yang X, Wang Y, Yu Y, Huang N, Li G, Li X, Wu JC, Yang S. Artificial intelligence in drug development. Nature medicine. 2025 Jan;31(1):45-59.

[3] Tan X, Cheng G, Ling MH. Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence. 2025 Jun 1;8:100355.

[4] Hanna MG, Pantanowitz L, Jackson B, Palmer O, Visweswaran S, Pantanowitz J, Deebajah M, Rashidi HH. Ethical and bias considerations in artificial intelligence/machine learning. Modern Pathology. 2025 Mar 1;38(3):100686.

[5] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..

[6] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.

[7] Han H, Shiwakoti RK, Jarvis R, Mordi C, Botchie D. Accounting and auditing with blockchain technology and artificial Intelligence: A literature review. International Journal of Accounting Information Systems. 2023 Mar 1;48:100598.

[8] Kumar Y, Koul A, Singla R, Ijaz MF. Artificial intelligence in disease diagnosis: a systematic literature review, synthesizing framework and future research agenda. Journal of ambient intelligence and humanized computing. 2023 Jul;14(7):8459-86.

[9] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.

[10] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.

[11] Mannuru NR, Shahriar S, Teel ZA, Wang T, Lund BD, Tijani S, Pohboon CO, Agbaji D, Alhassan J, Galley J, Kousari R. Artificial intelligence in developing countries: The impact of generative artificial intelligence (AI) technologies for development. Information development. 2025 Sep;41(3):1036-54.

[12] Tu T, Schaekermann M, Palepu A, Saab K, Freyberg J, Tanno R, Wang A, Li B, Amin M, Cheng Y, Vedadi E. Towards conversational diagnostic artificial intelligence. Nature. 2025 Apr 9:1-9.

[13] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.

[14] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.

[15] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.

[16] Ozmen Garibay O, Winslow B, Andolina S, Antona M, Bodenschatz A, Coursaris C, Falco G, Fiore SM, Garibay I, Grieman K, Havens JC. Six human-centered artificial intelligence grand challenges. International Journal of Human–Computer Interaction. 2023 Feb 7;39(3):391-437.
[17] De Gagne JC. The state of artificial intelligence in nursing education: Past, present, and future directions. International journal of environmental research and public health. 2023 Mar 10;20(6):4884.
[18] Chiu TK, Xia Q, Zhou X, Chai CS, Cheng M. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education: Artificial Intelligence. 2023 Jan 1;4:100118.
[19] De Vries A. The growing energy footprint of artificial intelligence. Joule. 2023 Oct 18;7(10):2191-4.
[20] Vrontis D, Christofi M, Pereira V, Tarba S, Makrides A, Trichina E. Artificial intelligence, robotics, advanced technologies and human resource management: a systematic review. Artificial intelligence and international HRM. 2023 May 22:172-201.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 2: Artificial Intelligence in Cyber Defense: Zero-Trust, Automation, and Autonomous Security

1. Introduction

The most secure organisations are afflicted with cyber threats; attack outcomes are severe to the victims and have a ripple effect on the global supply chain and responsible structures. Artificial Intelligence (AI) promises to provide affordable countermeasures to more sophisticated attackers who can exploit automation to use a greater number of attack mechanisms. Identity analytics based on roles form the foundation of an integrated zero-trust architecture helping to reduce insider risk. AI makes it possible to achieve the automation of decisions in the methods of cyber operations, even though their human control is also necessary to prevent Faraday cages caused by malware. The trend of autonomous security is becoming a reality; the machine speed defense is a worrying dream.

The SolarWinds and Colonial pipeline incidents in the recent past can be learnt. The fact that SolarWinds had an insecure supply chain of software shows that there is no organization that is capable of removing all the risks. An extreme example demonstrates that resilience is more important than prevention; AI facilitates the quick recovery of services and services. SolarWinds sheds light on the risk-reduction methods to be used in any key function: focus testing capabilities to improve; use red-teaming to test resilience; review architecture with the perspective of AI, automation, and operational coordination. The reaction to the Colonial Pipeline incident explains the role of exhaustive contingencies planning, testing, and additional funds in the enhancement of resilience.

2. Foundations of AI-Driven Cyber Defense

The research on AI-based cyber defense and evidence suggest the concepts of Zero-Trust Architecture (ZTA) and Identity Analytics (IA) as one of the essential exploratory factors, supported by Automation and Orchestration of cyber activities and, at the next level of maturity, Autonomous Security when the speed of computing is the central factor. There will be declared as the largest engine of "change and potential growth" of U.S. defense in the next decade and further on based on AI. It is important to keep investing in AI to enhance the resilience to dynamic adversaries. The AI-powered infrastructure should be able to generate the appropriate data, accelerate the time of detection and mitigation, and cascade them together when dealing with systemically significant infrastructures.

AI production-scale executable can be used to support more robust Zero Trust involving IA and contribute to the construction of defenses against more severe cyber threats. The AI investments focused on ZTA-based cybersecurity are directed by the public sector, and the funding adds to the policy, acquisitions, research, and the formation of the workforce. These initiatives find reflection in key partnerships with the private sector e.g., the Cyber-Resilient Electric Energy Delivery Cyber-System (CREDS) program such as market leaders announcing that they provide technology and services to

critical infrastructures. However, it is agreeable that still active AI is not fully implemented in such productions yet.

2.1. Threat Detection and Response with AI

There are various institutions that have automated threat detection and response when carrying out cybersecurity. However, it may use conventional rule-based technologies instead of advanced machine learning (ML) and artificial intelligence (AI). The rise in the use of AI and ML has fundamentally changed how systems detect, respond to, and mitigate threats. It is, however, critical to understand how AI and ML are being implemented in this foundational building block of AI-driven or AI-enabled cyber defense when the technology is relatively new or still immature. Understanding the factors that make AI or ML maturation in these threat detection and response deployment areas is essential for other AI and ML deployment areas. These deployment areas can range from cloud AI, such as Microsoft Cloud or Amazon Web Services, to automated autonomous responses at machine speed.

Nevertheless, there have been significant investments in ML and AI for threat detection and response systems in both sectors. AI and ML capable solutions are growing by more than 25 percent annually and appear to be the foundation of a new AI-driven cyber defense architecture that is already going beyond these areas and increasingly toward AI-enabled cyber defense[1-2]. The majority of these solutions involve both threat detection and response with increasing use of autonomous responses. Defense policies, however, often focus only on the broadening of threats, detection, and or response capabilities based on conventional rule-based systems. The transformative nature of AI/ML in these sense-and-respond building blocks of future AI-driven cyber defense is just as much if not more important for cyber defense and society as a whole as are the emerging evolutionary AI/ML solutions more commonly deployed in society. Yet the nation must achieve a much higher level of maturity in these areas to be fully effective as another key enabling foundation.

2.2. Zero-Trust Architecture and Identity Analytics

All information technology (IT) assets and offered services are presumed untrustworthy, along with any entity seeking access to IT resources. AI-powered identity analytics, detection and response ability, SIEM, and behavioral analysis processes enforce rules managing access and other interaction with IT-based services and resources in real time and change with time to provide better speed and accuracy. Machine learning models can establish a pattern of baseline behavior, which determines points of authentication, possibilities of access, and similar patterns. Normal behavior baselines generated by machine learning make continuous comparisons between activity of the user and the anticipated pattern of identity use with references to machine identity models, process models, behavior models, user account, and device profiles.

An identity, identity usage and derived attributes such as risk, dynamic context and secure willingness to transact randomly that can cover the risk surface area integrates real-time anomaly monitoring against communication patterns and other entities that are expected to act are involved in this. Individual identity monitoring encompasses risk surfaces against foreign transacting entities regardless of the perceived benign nature of the specific client. Groups with continuously modeled expected behavior manifest risks invalidated during testing to guarantee maximum signature protection building trust in the validating vector signature against all other groups. Characteristics of the transaction and testing also examine transactions for important probability pairing and cross-support.

2.3. Automation and Orchestration in Cyber Operations

Through the speed of information retrieval, discovery and correlation, AI can identify patterns or links in unlimited amounts of data and advance these results to levels of threats, at the same time eliminating irrelevant baseless signals. This rapid evaluation capacity signifies AI's major impact in cyber defense, especially in its provision of new detection and mitigation measures for the APT paradigms that have previously plagued the cyber arena. Shadow IT and APT rapidly adapt to offense-plus-defense game-theory paradigms for a given target or goal, and while behavior-based detection aids cybersecurity defense, exploiting unrealistic assumptions on the accuracy of such detections limits utility. AI, however, holds the prospects of significantly reducing complete automation demands in response to APT and extending on a dropped assumption of zero-feedback loops in adversarial models using game theory.

In small spaces of interest, AI-directed Learning Automata (LA) can still be able to model and deploy these counterparts and protectors with less training of the latter. The ultimate challenge of cyber defenders is to automate the entire range of the detection, response and recovery operations and reduce dependency on human participation as much as possible. With these considerations in mind, functionally automating and orchestrating cyber operations at machine speed emerging as the strategy for cyber defense within the rapidly evolving APT game-theoretic paradigm. Such developments also include the establishment of a machine-speed cyber defense system capable of cross-enterprise response-and-recovery orchestration on behalf of all stakeholders in critical infrastructure services, followed by creation of a machine-speed complete-service in-circuit cyber-attack-spectrum detection and response exploitation.

2.4. Autonomous Security and Machine-Speed Defense

The very close idea to the concept of automation is an autonomous security which is offered by systems and attacks a certain type of threats without human interaction.. Autonomous systems can offer timely response for basic categories of incidents, such as denial-of-service attacks or easy-to-spot malware infections[2-4]. The resources such as Cloudflare or Akamai cover big infrastructures against the denial of service attacks that are easily noticed and shields that require too long to use. In this case, recovering service of the victims is just a case of routing and redirecting. The more these routes are previously laid the quicker the victims heal. These responses fall under non-centralized technical automation that is attribute-based.

The most critical part of AI-based cyber defense is automation which involves machine-speed detecting of demands and automatic responding to an incident. These components must be computerized, choreographed and demonstrated living solutions. Telegraphs, railroads, trucks, and boats are such motorways in this automated city of living, the motorways of demand detection, routing and supply. There are also distinctions between formal cyber demand, things, and people that require central IT response, to the larger set of demand detection in the physical-world that are excluded. Enterprises need a control tower that can detect that an airplane is falling and inform central IT.

3. Government and Industry Implementations

The argument used in support of the idea of accelerating the implementation of AI-based cyber defense includes the consideration of the real applications that were realized in the public and within the corporate spheres. The U.S. Department of Defense has the highest level of interest in the use of AI in cyber operations among the entities of the public sector under analysis. The literature review presented earlier

highlights the persistent focus on public sector adoption within academia, but industry implementation receives equivalent attention in the form of critical infrastructure protection. Although AI-based defense strategies remain nascent, their implementation is nonetheless maturing.

In the United States, the Executive Office of the President recognizes the effort of the private sector in supporting the government's zero-trust goals and accelerating the transition of AI- and machine-learning-enabled technology The lessons gained by AI by government partners and the private sector suggest that AI does not provide a silver bullet, it is simply a tool that may be used to spot threats. Meanwhile, there have not been simple definitions of AI in the cyber context. There is also work to do in order to make sure that AI deployment will not negatively impact the security posture of organizations or internal controls as well as prevent relying on AI externally to facilitate the use of a zero-trust strategy. AI also must be resistant to the biases of those that implement and support it.

3.1. Public Sector Deployments and Policy Context

The use of AI-based capabilities by the federal government and dependence on developments in the private sector create one of the main frameworks of adoption. The AI is also involved in critical infrastructure protection and administrative support operations, and initiatives, including the AI Strategy to the Federal Government, the National Security Commission on Artificial Intelligence report and the Department of Defense AI Strategy are moving forward. Cybersecurity, disease forecasting and prevention, climate adaptation, recruiting and retention, and human-machine teaming are some of the prominent areas of AI investments and pilot projects.

The National Cyber Security Strategy 2023 underlines that the cyberspace should become a safe and reliable environment. According to it, with backing by the private sector, the United States needs a higher sense of security in the cyberspace in all the spheres of the society, such as government, business, civil, or individual. It involves speed in responsibility to cyber incidents and vulnerability; introduction of best practices with regard to cybersecurity; and further reduction in underlying risk compelled by organizations to regard their own security and its impact as their responsibility." Given that "that the private sector owns and operates most critical infrastructure and controls the technologies and digital world that supports government, foreign policy, and a vast array of critical functions," AI is considered essential to enable the government's role in protecting critical infrastructure via partnership with industry. National security enterprises (NSE) need to develop and implement capabilities to respond to the alignment of the strategic intent and AI assists the Cybersecurity and Infrastructure Security Agency in the likeness of providing prompt alerts to elicit warning.

3.2. Critical Infrastructure and Private Sector Adoption

Cyber threat vulnerability to critical infrastructure of transportation, finance, and energy sectors poses a threat to the national security system, economic destabilization, and physical insecurity[5-6]. Government investments in cybersecurity are accelerated by the bipartisan Executive Order on Strengthening the Cybersecurity of Federal Networks and Critical Infrastructure in Win spurred by the SolarWinds incident and the ransomware attack on Colonial Pipeline.

Critical infrastructure in the United States represents 20% of the annual gross national product; the critical infrastructure sectors are as follows: • Chemical • Commercial Facilities • Communications • Critical Manufacturing • Dams • Defense Industrial Base • Emergency Services • Energy • Financial Services •

Food and Agriculture • Government Facilities • Healthcare and Public Health • Information Technology • Nuclear Reactors, Materials, and Waste • Transportation Systems • Water and Wastewater Systems

The Government Accountability Office (GAO) assigned a high-risk designation to the programs related to cybersecurity in critical infrastructure. Despite the existence of Federal Information Security Modernization Act metrics, public-private partnerships with information sharing frameworks, investment incentives, and grants, public-private partnerships, and investment incentives and grants, empirical data show that the implementation of cybersecurity controls still lags behind within the private sector. The recognition of AI-enabled security resources in the private sector is illustrated by the growth of the sector with respect to firewall, anti-malware, SIEM, management and testing tools, and identity and access management systems.

4. Case Studies and Lessons Learned

Even though the recent events of artificial intelligence-enabled cyber defense have been increasing in interest and investment, each of the recent events of such a nature has not been mitigated or detected in real time. The analysis of the famous SolarWinds and Colonial Pipeline attacks, however, provides some good and bad data. A cross-case study grounded on these two cases, as well as a third case that is less known to many people, supplies insight to make successful application and management of AI in cybersecurity[7,8]. New attack patterns, effective filtering of false positives, better incident response time, risk-identifying important assets, and enhanced vulnerability discovery as well as patching are also some of the problems that machine learning solves in cyber defense.

This 2020 cyber intrusion targeting SolarWinds Corporation and its customers such as the various government departments, several states of the United States, and over 100 companies acted as one of the dramatic wake-up calls in the ten-year war battle against cyber threats. Such high-profile and long-term exposure brought the need for more efficient threat detection, dynamic response to new attack patterns, and better protection of government agencies and companies to the top priorities of national cybersecurity strategy. Large-scale investment in AI-enabled cyberdefense capabilities followed. Reliable data, sound algorithms, and sufficient processing power are key enablers for machine learning in cyber defense.

4.1. SolarWinds Incident: AI Roles and Limitations

Even though the recent events of artificial intelligence-enabled cyber defense have been increasing in interest and investment, each of the recent events of such a nature has not been mitigated or detected in real time. The analysis of the famous SolarWinds and Colonial Pipeline attacks, however, provides some good and bad data. A cross-case study grounded on these two cases, as well as a third case that is less known to many people, supplies insight to make successful application and management of AI in cybersecurity. New attack patterns, effective filtering of false positives, better incident response time, risk-identifying important assets, and enhanced vulnerability discovery as well as patching are also some of the problems that machine learning solves in cyber defense.

This 2020 cyber intrusion targeting SolarWinds Corporation and its customers such as the various government departments, several states of the United States, and over 100 companies acted as one of the dramatic wake-up calls in the ten-year war battle against cyber threats.. These tensions were highlighted in SolarWinds' 2020 massive cyber breach. Although SolarWinds is rightly cited as a real-world example of why AI is needed in the first place, the intrusion was undetected for months and did not implicate AI in its operation. Lessons learned reveal how AI was too little or too late. Telecommunications, cloud service,

and cyber security providers have recognized the need for a ready and comprehensive operational response to any major threat or exposure at any other provider.

Mobile user identity is being transferred to number of transport-layer command and control protocol to an automated, real-time and dynamically adjusted user identity attribution, such as user and device behavior learning.. Often referred to as identification analytics. Identity analytics attempts to streamline detection of authentication compromise, device and user impersonation, and user access privileges out of line with typical behavior. Rather than creating access control policies, the challenge is to learn and make decisions automatically by means of AI tools, and be concerned with the integrity of the connect surface. IPs, AV, and firewalls are combined and improved by improving data quality and context. Compared to traditional AV signatures, models are trained to respond to a composite of anomalies, as opposed to univariate detection of a single condition- multivariate validation- without necessarily beginning to detect individually at speed.

Three key weaknesses in SolarWinds' cyber defense strengthen the argument for integrating AI into a zero-trust architecture and identity-scale analytics: lack of threat blending, detection quality, and human process automation. A combination of detecting various indicators types, including network traffic analysis, endpoint behavior modeling, and detection by MES instead of the traditional AV signatures enhances the quality of the detection in general and minimizes the chances of evasion by the adversary. The capability to identify application-level malware as well as malware initiation on application servers is beneficial to bridge the gaps of the traditional internal network monitoring. The playbook execution and automation is needed, especially triggered by the attacks directed to threat monitoring and threat defenses features.

4.2. Colonial Pipeline Contigation: AI-Enabled Resilience

The speed of the detection, response, and recovery of AI-enabled systems can be emphasized by the fact that AI systems helped in detecting and acting upon the attack on the Colonial Pipeline and resuming the functionality thereof. Based on the usage of the vulnerability that leveraged a trusted third-party vendor, the attack brought down the functionality of the pipeline resulting in gasoline shortages and a subsequent increase in the fuel price[9-12]. Hackers quickly purchased ransomware and remotely accessed the target infrastructure and installed malware to hide the intrusion. Subsequently, the ransomware was executed in the operational environment, taking the pipeline systems offline. AI played a key role in expediting incident mitigation processes and in automating recovery from the situational rest state.

AI-enabled technologies permitted the rapid turnaround of the incident by anticipating and orchestrating the traffic-load redistribution to the unaffected segments of the pipeline. Every segment's volume demand was inferring from condition indicators. Dynamic descriptors of traffic volumes for each segment exiting the load centers to assure robustness to the numbers tracking performed automated decision-making and control while speeding the offensive and remediation campaign. Detection of attacks by trusted third-party vendors' terminals at later stage and signature-based protectiveness had also been automated. Automation across different resolution layers permitted the reduction of mean time to detect and to respond. A combination of minor and moderate improvements reversed the trending towards long responses to major incidents.

4.3. Cross-Case Analysis and Mitigation Strategies

A comparative examination of the SolarWinds and Colonial Pipeline incidents helps identify opportunities that could alert security decision-makers to future incidents or increase defense system resiliency and survivability. AI technologies certainly provide effectiveness and added resilience, particularly for timely detection and minimizing damage from immediately propagating vulnerabilities such as those exploited in the SolarWinds breach. But AI-in-the-loop supporting de facto technology adoption matters could also reveal why the vulnerabilities had went undetected when they could have been discovered quickly and why thus the AI support with detection-là-for-travel fonctionnality was not deployed.

A path for AI adoption-in-assurance for safety of a real cross-boundary operation could appear through the mix-support evidence of AI effect on the rainwater filtration system set in Taiwan in case part of an incidebts road map plan for flooding risk.

5. Challenges, Risks, and Ethical Considerations

There are no risks and difficulties free regarding AI-driven cyber defense. The software relies on the quality and unbiased data, which is hard to be attained. During training and operation, privacy violations are still an issue, bias, discrimination and audit are also a problem. Intelligent defense systems are subject to the manipulation of adversarial attacks which can cause intelligent systems to falsely label malicious actions as harmless [7,13-15]. The future of these systems lies in the ability to debate their ethical overlay in a sound and comprehensive manner and what consequences it may have. These discussions must address the areas of salience to public and private sector decision-makers constructing new or augmenting existing systems.

Data Quality, Bias, and Adversarial Attacks Although there is increasing evidence around the efficacy of AI in cyber defense, more development and testing is required of systems. The quality of training data is a factor of AI decision-making: garbage in, garbage out. In addition, models that are trained on data that has human bias (e.g. trained security analysts indicating a certain type of data exfiltration) can be biased. Moreover, cyber enemies continuously use different types of tactics, techniques and procedures (TTPs) in an attempt to breach systems and networks throughout the kill chain and avoid detection systems at any stage. These adversaries have added another layer to their TTP arsenal, using generated adversarial examples to distort and deceive classifiers, and fool complex detection systems based on varying forms of deep learning. These examples can be identified as non-true to various classifiers and measures to pose as perturbations of rightly classified AI inputs or very spatially clustered poisoning points.

5.1. Data Quality, Bias, and Adversarial Attacks

Advances in AI for security depend critically on the integrity and quality of the data the systems ingest. The participation of criminal, hacker, or hacktivist actors, who pursued their own/foreign policy goals, such as downing the colonial pipeline and deploying Marley, an AI-assisted troop, remains alarming. In various incident and attack patterns AI exhibited fragile accuracy. For example, on May 26, 2021, security for Decision Making Engine (DME) detected, classified, and responded to dozens of malware, but only one was a “false positive” (Casa, Fong, Hamada, Kim, & Wu 2021); malicious code Arkime had never been previously seen and was changed every hour to evade threat detection. In other cases, such as the use of Deepfake technologies in scams and the AI-INDWEST/MarvoBots campaign, the operationalization automation-symbiosis pattern neural networks were able to mask the appearance of

their true actors—fraudsters or ESTs behind the curtain. Therefore, Mitigation-Misuse Governance becomes critical because adversaries can either attack the integrity of these systems or apply reverse-engineering techniques to copious data sets obtained from the operational use of supervised Machine Learning algorithms, generating adversarial samples and fooling networks, but Neural Networks are not the only module that can be attacked.

The race between information security and information security attacks can express inequality for both sides, jeopardizing the equilibrium. On one side, it is trying to put more capabilities on control: the detection of specific malware patterns on specific days for rapid mitigation and concentration of human analysis on other situations; on the other side, it is totally exploring capabilities (completeness and performance) on synergetic aspects.

5.2. Privacy, Compliance, and Trust in AI Systems

Privacy is a critical security issue, which has to be strongly regulated and guard railed by business. Regardless of an ethical developmental and usage promise, the lack of trust is there with its bad actors and the necessity of solid systems. The implementation of AI privacy is aimed at reduction of the risks of data retention, use, and misuse. The models are shielded by the AI systems that ingest aggregated and anonymized information and with the aid of federated learning and differential privacy approaches. The presence of compliance is also a problem of concern: the reverse engineering of an AI model may disclose the training data; the proprietary knowledge may be left exposed during test queries. Those risks are addressed by the AI regulations stipulated in the draft EU legislation.

A detailed data governance scheme examines all data types regarding regulatory and contractual requirements, business value and risk, data reduction, privacy influence, access controls, life cycle administration, security imperatives and automated and relentless surveillance. Organizing more supervision, automation, and information efficiency, organizations will make AI solutions more trusted as they make their work more efficient [9,16-18]. The ethical responsibility of AI systems can also be distributed to create confidence among the populace and the consumers. Organizations who strive to explain the systems of AI necessarily have to strike a balance between inclination and simplification such that the conclusions and directions of the model can be followed without being lost through too much simplification.

The quality of the data is key, and it is paramount, particularly when it comes to zero-trust environments. Learning the lessons of the great events may help to create an appropriate base, nevertheless, it is possible to hide weaknesses and introduce new issues. Data reliability is further curtailed by unstructured data sets without any knowledge, sloppy introduction of unproven learning procedures, and lack of communication between different departments. The lack of proper consideration regarding the quality of AI systems may also cause an extreme quality of loss in data and, consequently, massive disruption.

5.3. Governance, Accountability, and Explainability

Although AI systems are expected to alleviate the roles and burdens of human operators and detect novel threats, they are not immune to external manipulations[2,19-20]. Adversarial attacks against machine learning (ML) models are especially complex to identify. Furthermore, even minor perturbations may achieve significant successes: different nouns such as toad and cat were previously classified using a similar image but with mere rotation, size reduction, and a single point perturbation. As such, and as

discussed in Section 5.1, genuine and labeled data are vital for AI training and regularization against adversarial circumstances.

AI is typically viewed as a black-box technique: a logical reasoning mechanism for fusing low-level signals into high-level alerts with little understanding of operating significance. Such a scenario may impair user trust; hence security, intelligence, and military agents still prefer rule-based systems such as Snort and Yara due to their intelligence and effectiveness. Explainable AI (XAI) aims to articulate detection reasoning and alerts by quantifying sensor importance, explainable model capabilities, and the underlying operations of black-box detectors. XAI may subsequently enhance understanding of abnormal activities in addition to facilitating causal investigation and enabling confidence under uncertainty. Such notiore expect a good design of AI Frameworks through Governance and Projects.

Governance defines accountability roles and procedures and enables management of technology and business risks. The AI governance frameworks have the tendency to focus on business initiatives at the expense of explicit AI governance requirements. Nevertheless, AI governance provisions serve as a critical need, particularly in the quest to exploit the mature AI-based features of the digital transformation of AI-based application security. The aim of such efforts is to ensure that the quality, reliability, compliance with the regulators, and practical performance of the AI-based smart product design and implementation will be met prior to the implementation.. Sub-areas of AI governance address the security, privacy, and resilience of the design and operation of the AI-based product with respect to the adversarial-economic-based-actuation of malware, as also discussed in preceding sections.

6. Roadmap for Implementation and Maturity

To reach the target of AI-based cyber defense, stringent use of automation, vulnerability detection, case testing and responsiveness at speeds of machines will be required. A maturity model evaluates the paths of the government agencies and the private sector initiatives, technical trends, and analyzes the operationalization of the aspects of zero trust. It implies pragmatically defined scope planning application in the application of AI in defense against foreign malicious cyber attacks and provides recommendations on successful usage.

While AI operates at machine speeds and is extensively used in cyber-technology product governance, the availability of operational AI-driven defenses at the same speed is limited Historically machine-speed reactiveness to age-fashioned cyber operations e.g. taking advantage of weaknesses in publicly provided services, haphazard access and exfiltration of data, sideways movement and ransomware implementation has not been achieved [9,21-23]. The environment of deployment of that capability is one based on the development of exponential capability. It is not enough to power the already existing machine-speed defensive mechanisms; just as the offensive abilities of organizations have evolved into the AI-enabled automation, machine-speed defense must also have the AI-enabled automation its equivalent. These areas have maturity based in the development of training data that is constantly updated to effect supervised and semi-supervised learning.

6.1. Reference Architectures for AI-Driven Defense

Clearing House for AI-Enabled Cyber Operations Seven design summary architecture outlines the basic operations and levels of operation of cyber operations. Infrastructure Backing the deployment of AI through operations enge Hybrid Threat-Assessment Framework assesses whether an incident ought to attract an AI-tolerant of over AI level cyber capability reaction.

Concept-and-technology road map Reference architectures of AI-enabled cyber defense assist in managing developments of threat-actor modeling and data-validation testing to reach better results. Enterprise compliance to the concept-and-technology road map and enabling architectures is discussed: Foundations of AI-Driven Defense against Cyber Loss, Espionage, and Disruption, Publication Series M-2022-01-005-006. The warning Security-Engineering Lattice assists in prioritizing the pledges to protect the fragile systems, their sabotage, disruption and espionage within the minimum budget allocation.

6.2. Metrics, Testing, and Validation Frameworks

It will be essential to provide specific measures that will evaluate the safety functionality and the effectiveness of the AI-based defenses in order to create confidence in the possible users.. It enables organizations to determine which services they want to implement in an AI-driven Cyber Defense approach and gauge the quality of future deployments. Such metrics also facilitate testing and validation of deployed systems [24-26]. A suitable framework is needed to measure how different systems and implementations match their stated requirement in terms of completion and output.

A key consideration is how well AI-based capability deals with a particular type of adversary, making the argument of which set of metrics or test scenarios is least biased in favor of AI solutions a challenging task. Another consideration is that algorithms are often designed to follow a particular area of interest, such as object recognition. Clearly defining the requirements for deployment helps reduce bias by helping testers to remain focused on the area which is claimed to be secured by the AI implementation.The definition of metrics, test scenarios, sets of requirements, and an overall test and validation framework for the implementation of AI in Cyber Defense should, to the greatest extent possible, seek to utilize the tools and techniques of the adversarial community to validate AI capability.

The well-known saying of attack is the best defense explains the urge to have the necessary resources to penetrate a defense. Another way of derivation is to come up with a testing and validation functionality that is as resilient to defense mechanisms as defense mechanisms claim to be to an attack.

6.3. Operationalization of Zero-Trust with AI

The artificial intelligence (AI) technologies can be significant in operationalization of the Zero-Trust principles being applied through increased identity analytics, higher detection rates of threats, fewer false positives, and efficient orchestration of the parameters to have minimum influence on the user experience. Nevertheless, one of the principles of Zero Trust, the All-internal Resources Threat Detection, is currently difficult to be fulfilled by AI, as it can be seen in the case of SolarWinds mentioned above. The quality of the data is to be enhanced to prevent the cases of the introduction of false positives in detection pipelines. Furthermore, collaboration and data sharing between organizations is either limited or negatively affected when redundant alerts needed to be filtered [8,27]. The need for malicious attack frameworks and associated threat feeds that offer good coverage not just at the network but at all levels of abstraction—which provided a strong basis for mitigating the Colonial Pipeline ransom incident—is also evident.

Consequently, the path to AI for operationalizing Zero Trust requires zero-trust enabling organizations to create an initial AI capability roadmap outlining which AI-enabled functions and components need to be delivered in the short, medium, and long term. Progress will provide the necessary defensive value foundation coupled with increased maturity and confidence to support further engagement and utilization of AI in other areas, within and outside the organization. Another recent use case framework of AI-Zero Trust noted 32 AI-powered use cases aligned with all nine of the Zero-Trust principles, and how to

operationalize Zero Trust using AI. To provide a general perspective, the use cases include identity, credential and access management; device security; network security; application security; data security; security analytics; security monitoring; security operations and third-party risk management. Among the use cases, 21 are cases which have been adopted by organizations which are at different maturity levels and they include strong, striving and nascent.

7. Conclusion

Since cyber threats are becoming more numerous and complex, emerging defensive capabilities based on artificial intelligence (AI) are usually considered as essential to facilitate the improvement. Although the degree of maturity of AI solutions targeting cyber defense is still not even, numerous government agencies and organizations across sectors have begun to consider using AI to provide value to improved and faster threat detection, identity verification, response to incidences, secure automation, and other functionalities. But, similarly to any technology, AI is associated with risks, challenges, and ethics that should be implemented and overcome to conduct safe implementation and functioning.

A roadmap would give guidance to the exploration and adoption of AI as the inferential measure towards cyber defense. A reference architecture to help users adopt the roadmap in the public sector is backed by the principles of a reference architecture and is measured and tested with an AI-enabled capabilities testing and validation framework.. The reference architecture also facilitates the operationalization of AI in support of a zero-trust security strategy. Such contributions are aimed at bridging the gaps in the implementation and comprehending the meaning of implementing AI in cyber defense. These actions would enhance loyalty and maturity, be more resilient, respond more effectively to foes, and promote the use of AI in cyber activities safely and efficiently.

References

[1] Korzynski P, Mazurek G, Altmann A, Ejdys J, Kazlauskaite R, Paliszkiewicz J, Wach K, Ziemba E. Generative artificial intelligence as a new context for management theories: analysis of ChatGPT. Central European Management Journal. 2023 Mar 28;31(1):3-13.

[2] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.

[3] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.

[4] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.

[5] Zha D, Bhat ZP, Lai KH, Yang F, Jiang Z, Zhong S, Hu X. Data-centric artificial intelligence: A survey. ACM Computing Surveys. 2025 Jan 24;57(5):1-42.

[6] Papagiannidis E, Mikalef P, Conboy K. Responsible artificial intelligence governance: A review and research framework. The Journal of Strategic Information Systems. 2025 Jun 1;34(2):101885.

[7] Shivadekar S, Mangalagiri J, Nguyen P, Chapman D, Halem M, Gite R. An intelligent parallel distributed streaming framework for near real-time science sensors and high-resolution medical images. In50th International conference on parallel processing workshop 2021 Aug 9 (pp. 1-9).
[8] Yim IH, Su J. Artificial intelligence (AI) learning tools in K-12 education: A scoping review. Journal of Computers in Education. 2025 Mar;12(1):93-131.
[9] Gursoy D, Cai R. Artificial intelligence: an overview of research trends and future directions. International Journal of Contemporary Hospitality Management. 2025 Jan 2;37(1):1-7.
[10] Sengar SS, Hasan AB, Kumar S, Carroll F. Generative artificial intelligence: a systematic review and applications. Multimedia Tools and Applications. 2025 Jun;84(21):23661-700.
[11] Ferrara E. Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies. Sci. 2024 Mar;6(1):3.
[12] Panda S. Secure Access Management in Serverless Computing Through Blockchain Integration.
[13] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.
[14] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[15] Han R, Acosta JN, Shakeri Z, Ioannidis JP, Topol EJ, Rajpurkar P. Randomised controlled trials evaluating artificial intelligence in clinical practice: a scoping review. The lancet digital health. 2024 May 1;6(5):e367-73.
[16] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.
[17] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.
[18] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.
[19] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.
[20] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[21] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[22] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.
[23] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..
[24] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.
[25] Samala AD, Rawas S, Wang T, Reed JM, Kim J, Howard NJ, Ertz M. Unveiling the landscape of generative artificial intelligence in education: a comprehensive taxonomy of applications, challenges, and future prospects. Education and Information Technologies. 2025 Feb;30(3):3239-78.
[26] Alwaqdani M. Investigating teachers' perceptions of artificial intelligence tools in education: potential and difficulties. Education and Information Technologies. 2025 Feb;30(3):2737-55.
[27] Cukurova M. The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. British Journal of Educational Technology. 2025 Mar;56(2):469-88.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 3: Digital-Twin Based Cybersecurity for Critical Infrastructure

1. Introduction

Here, I delineate digital twins in the context of the critical infrastructure components and systems within a particular region; in this context, 3 important questions are posed: 1) how are digital twins defined; 2) what predictive and early-warning capabilities can be available; and 3) what must be given particular attention when it comes to data governance, ethical, and risk management?

This centers on the power grids, pipelines, and healthcare systems in a large country, and water, transport, and logistics systems are also considered in this regard. The prediction and early warning features of such infrastructures are crucial when digital twins are employed, and the problem of availability, access control, privacy, safety, and risk management have an impact on the governance of data.

2. Digital twins in critical infrastructure

Digital twins for cybersecurity in critical infrastructure. Many national and global critical infrastructure sectors are adopting the concept of digital twins. With a broad definition, a digital twin is a model of a physical object, a collection of physical objects, a full physical system, or an ecosystem. The digital twins can be important in cybersecurity of critical systems. Digital twins represent power grids that make use of SCADA telemetry of Smart-DC substations. The pipelines have digital twins that reflect the intricate topology of transport networks as well as the cyber-physical sensors which offer information regarding the status and integrity of the assets and flows. Digital twins of a healthcare delivery system include the modeling of patient flow, real-time telemetry of clinical equipment as well as governance dimensions linked to the confidentiality of sensitive patient data. Further studies are required to experiment with further digital-twin cybersecurity concepts and architecture on a larger scale in different infrastructure areas on top of the initial use cases mentioned.

Banks, insurance companies, and other infrastructure cyber-risk owners have telemetric data-pipes that have sub-second latriminal telemetry in real time. Digital twins of China’s power grid exhibit shown components such as transmission lines, substations, substations network, electrical connectors, SCADA telemetry data, weather data, and electricity market data. Power grid management relies on such telemetry data for real-time monitoring and on-line control of the grid operation, which are very important for thread prevention and risk mitigation [1-2]. A typical telemetry system is composed of SCADA-Supervisory Control and Data Acquisition system/protocols, ICCP- Inter Control Center Protocols, NGDMS-Next Generation Data Model Server, and Smart-DC. SCADA-Supervisory Control and Data Acquisition system/Protocol data acquisition from several remote terminals and provide integrated data to

Control Center. The Control Center takes the data and apply analytic model to discover hidden information to support operator's decision making.

2.1. Digital twins of power grids

Power grids are key assets of modern society and a priority target for cyber and physical attacks. Virtual representations of power grids can improve grid-grid interactions and aid four essential functions, namely, data-driven prediction of models, predictive models, predicting of rare events, real-time threat prediction, and modeling of failure propagation. The grid digital twins are supposed to be developed based on the multi-layer model framework that indicates the grid topology, assets, physical behavior, real time data feeds, and dependencies on external systems. The development of such digital twins has to take into account the fact such common cyber-physical attack is silent and requires anomaly detection.

Digital twins of power-grids and energy-systems have become frameworks of cyber-physical-modeling that have the potential to improve the comprehension of dynamics in power-grids, and assist in completion of essential work like data-driven model prediction, predictive simulation of rare events, real-time threat detection, and failure propagation modelling. The important aspects associated with their design are the adequate modularity and completeness of the constituent operational models, heterogeneity and accuracy of input data feeds, and proper altruism of the links with external systems, in particular, water.

2.2. Digital twins of pipelines

Digital twins of pipelines

Digital pipeline twins are sophisticated physical objects because they monitor the working of pipelines. It is their fidelity because they include numerous sources of data, such as the data on telemetry sensors, which are used to determine the fundamental measurements of fluid properties and flow, at near positions, modelling software, that is used to calculate loads along the pipe based on real-time operational data and predicts any abnormal propagations, and automated tools, which derive new information on periodic inline inspection using intelligent pigs [3-5]. Like the case of the twin models in power grids, in pipeline digital twin, there is also a need to have predictive models on the risk of occurrence of an incident. The estimation of propagation of any form of events i.e., the spills events, land instability events, unauthorised events, ruptures and releases into the freshwater system give a complete picture of the risk to the network, and inform the decision support process in the event of a crisis.

The pipeline management systems apply real-time AI to the system. The risk-prediction and fragility models provide the digital twin with information and assist a quality-and-freshness-data judgment, and outline the system of a neural-network framework of the telemetred risk-detection in the typical operation state. Detection is performed by a three-layer tensor-decomposition model, for which the training dataset comprises a great number of operational situations and won the Silver Medal at the DICE-2020 competition.

2.3. Digital twins of health systems

Both the digital twins of patients undergoing surgery in hospitals and the digital twins of the hospitals encompassing operating rooms, wards, laboratories, and support areas are digital twins of health systems. The digital twins of patients undergoing surgery are simplistic. The emergency flow of a group of patients and the flow of patients undergoing surgery are coupled and managed. The medical devices containing

sensors on a patient's body act as supply chain 1 and with a mesh architecture with an alarm. The forecasting (flood/cyber) and decision support help the model run the health system fast and seamless without divergence. The digital twins of hospitals and other healthcare institutions replicate the floor plan and patient flow. These structures seldom affect process models but help while planning and implementing new requirements.

Healthcare systems are governed by different types of institutions like hospitals, laboratories, and support organizations, and health devices. The digital twins model the link between the institutions, the different devices in those institutions, and the operating procedures in place. The institutions are locations of greater specialization with respect to flow and are supplied by the other institutions by sending or receiving materials. The set of devices are the devices that collect parameters related to health safety and the support procedures are alarms with respect to the patient check and flow support. The digital twins are useful in testing the continuity of the functioning of the system and in modifying the operating rules with the patients put in a special condition.

3. Predictive analytics for cybersecurity

One approach to anomaly detection extends the tensor decomposition technique to higher-order data structures and to wider classes of anomaly types. Tensor decomposition can also support more predictive cybersecurity applications. Tensor-based relational forecasting can predict future values of nearly any variable, and may be exploited to set up a warning system. When information about the entire system is available, synthetic stress tests can help detect properties that influence resilience. For more focused decision support, a planning module estimates the worst distribution of notional hazards that the system can withstand. Finally, decision arguments can be generated for practical resilience measures, using additional models of the considered system if necessary.

Higher-order time-related sensors can be modelled not as an ordinary time series but rather in a three-dimensional tensor whose index set corresponds to the discrete time marks and, normally, to the geographic location. There may also be spatial warning signals provided the extra dimension is of a reaction relevant computation on surrogate sensors. The third dimension may theoretically accommodate any other categorization such as the nature of an anomaly. But in practice, it is not essential to discuss the cause of such anomalies, but to identify the potential riskiness of spreading across the network within a short time, with the detector detectors used in it, and the aggregate history of the observations.

3.1. Tensor decomposition for anomaly detection

According to estimates, in 2019, 152.70 million records containing one or more pieces of personally identifiable information were lost or stolen. In 2020, COVID-19 laid bare many of the healthcare systems weaknesses, and hackers used this juncture as an opportunity to exploit vulnerabilities and weaknesses in the health-related systems via phishing, data theft, Internet abuse, extortion, malware, and Denial of Service (DoS) attacks, among others. As such, privacy must be given due consideration and be managed properly, considering user concerns about personal information handling. For example, in 2021, the shift from business as usual to remote work and the growing reliance on digital solutions have introduced a surge in the volume and complexity of cyberattacks [6-8].

Business and IT environments are advancing faster than businesses and their protection measures can keep up. The point is to make different stages connecting AI (artificial intelligence) adoption with timely reactions, real-time monitoring, and twinkling insights of a near future and taking steps before a bad thing

occurs as one way to solve this dysconnectivity.. In order to carry out an effective and timely real-time monitoring, it is vital to adopt a formal and comprehensive architecture that supports the major steps involved. The architecture needs to support the fast generation of the multiple data feeds from the monitored systems together with the timely merging, correlation, and determination of any relevant alarms. Any precursors that could damage assets must then be dealt with in a timely manner.

3.2. Forecasting and resilience planning

Digital twins and predictive analytics are used in forecasting and resilience planning, which simulates various contingencies and various other futures of critical infrastructure. Sophisticated scenario modelling and stress testing allow exposing risks and work out an effective risk-reducing strategy. The changing threats and risks will require a scenario-based planning on the preparations of different future situations. Cybersecurity controls have traditionally been validated through penetration testing, but significant failures continue to occur. Digital twins can assist in scenario modelling to assess resilience and test strategies for coping with unexpected hazards.

Digital twins enable logical modelling of services on critical infrastructure, which can then be subject to scenario modelling. By exposing different logical and physical elements to intelligent stress tests, test directeur digital twins can identify risks that are otherwise hidden. Combining the testing results with an impact-propagation model provides a reliable basis for decision-makers and senior operating personnel to make investment and operational decisions. To ensure independence, the work is best carried out separately from the operating organization and then shared in a way suitable to the audience. Before a stress test, the key management decision-makers convene with the simulation director to discuss an enclosing range of key alternate futures to test for.

4. Early-warning systems and real-time AI monitoring

Real-time cybersecurity monitoring architectures have to support continuously growing scales of data produced by the telemetry, alarm detection, threat intelligence, and mitigation systems. Whereas early-warning systems forecast the future to prevent the on-coming attacks or measure system robustness or extreme circumstances, real-time architectures rely on detections alerts along with fusing multilayer threat-intelligence feeds. These pipelines are best reacted to by automation, but the presence of a human-in-the-loop activity is essential to action [7,9-10].

The major needs to satisfy in order to detect threats on time are identified and used to influence the design of prototypes resulted. These were spread out to several critical infrastructures sectors, and were useful in securing pipelines, energy, water, healthcare systems, transport networks, and logistical systems- areas that weave attraction to hackers. Specific properties such as the origin and frequency of repayment of different feeds of intelligence, the anticipated delays of acquiring and processing data and procedures of alarms and managing facts are among the properties which will be considered in dealing with the information overload issue.

4.1. Architecture of real-time monitoring

Digital twins are real-time monitored by addressing the data pipelines, data processing engines, artificial intelligence and analytics, alarm management, and assets that transform raw data into output decision-support. Latency goals are dependent on the usage: data-based regression usually takes less than 2 minutes, rule based safety notification less than a minute and real-time anomaly detection a few seconds.

It may not be the case with other elements like training the AI model or decision-support tools application.

Early-warning and real-time AI monitoring is based on the digital twins that are constantly fed on classical telemetry, the data of the cyber-event logs, and the external data. ICSs, SCADAs, IoT devices, and cloud services are data sources that produce high-velocity data streams. Risk analysis, on the one hand, supplies information on risks and threats, and on the other hand, cloud service providers, malware feeds, and dark web monitoring data are added. Data which demand low latency (such as health, integrity, and security data of safety-critical alarm) pass through a decision-support or alarm-management layer. The rest of the data are directly fed to anomaly models and forecasting models. To guarantee low latency, then an architecturally layered design is adopted, such that the origins or destinations of each processing layer are specified based on acceptable limits of latency.

4.2. Threat intelligence integration

The integration of threat intelligence feeds between various feeds into a central database is more likely to make the Early warning systems and real-time AI based cybersecurity monitoring interact more effectively. An example of sources would be the social media, the dark web, as well as patriotic hacker groups, data obtained by the public sector (including CVEs) can be done in-house, whereas external data is offered by commercial feed [1,11-14]. The source may be any source provided that they present information in the machine-readable form precisely and at the opportune time. Alerts associated with specific threats are transformed into queries to be answered by public or private databases. Ultimately, the time taken to receive the required information must not become a source of exploitable risk. The fusion mechanisms determine reliability and credibility scores of individual feeds and sources of data enabling information to be weighted and then injected into a decision support system.

Timeliness is crucial, as one can react faster to the system of early-warning only if the speed of data processing is sufficiently high, as are the sufficiency and quality of the implications delivered. Handling the huge volumes of data produced by multiple threat feeds, is just as problematic as with the alarms of the early-warning system: not all possible cases must produce an alarm.. Hence, introduction of a taxonomy focused on the potential risk of a threat. Information that does not meet the minimum requirements does not require further consideration, but information received from highly trusted sources takes precedence and can trigger human-in-the-loop processes.

4.3. Alarm management and decision support

The system of alarm management, decision support systems allow responding to vulnerable services or assets that are at risk of being compromised in time. These systems are usually structured in the form of alarm clouds, which consists of different autonomous alarm streams in regards to the respective services, and these alarm streams are complemented by cross-domain alarms. An alarm management framework is meant to select false alarms, group correlated alarms and give priority to minimal and actual alarms. This is achieved through determining impact awareness-the ability of corporate units to determine the extent of the impact of possible service loss or degradation on the business purposes and safety.. Properly managed alarms provide operators with a clear and simple review of infrastructure threats and provide the security center with distilled and prioritized information.

The digital twin of critical infrastructure is trained to evaluate a broad list of alarms and offer an alarm prioritization mechanism. Several prioritization methods can be developed based on real data such as risk

analyses, heat maps, or simulation results. Alarm-response suggestions and procedures are continuously updated in order to assist the operator during the control of main infrastructure services powered by decision-support systems [13,15-17]. The operator's role and decisions remain crucial, informed by an updated complete picture and rationale.

5. Protecting water, energy, healthcare, and transport systems

The process of securing water, energy, healthcare, and transport infrastructures can be coordinated by cyberthreat-defence in order to prevent physical, digital, and human aspects attacks. The systematic evaluation of the assets and risks guarantees that protection against vulnerability and the risk done is commensurate.

Infrastructure Security of Water. Water-processing plants, supply-storage-distribution networks, and SCADA and Internet-of-Things, as well as corporate ticketing, need to be assessed risk-to-vulnerability in layers in order to test the worst-case breach scenarios. The analyses aid cyberhygiene planning and training on zero-day vulnerabilities on a long-term basis. Periodic test of incident-response plans with known risks by sticking and observing the quantifiable results of the responses would verify the preparedness of the actors to react in time.

Cybersecurity of the Energy System. The cybersecurity hardening of generation, transmission, distribution, and system-management systems against structured, unstructured, and proxy- ransomware attacks through the use of digital twins of the energy sector; extending to the resilience of the sector to disruption of supply of renewable-energy; and to guard against disrupted equipment throughout the energy- supply chain [18-20]. A tiered alarm-management system is a system that aligns the alarm sensitivity to the system capacity, that the high threats are detected in real-time to be subject to human intervention, whereas the low-level aberrations will resort to investigations to analytic intelligence.

Security in the Healthcare Network and Devices. Healthcare network, associated devices, and implantable medical device cybersecurity frameworks put risk assessment and risk management ranking on the values of reliance on networked devices, schedule of devices patched, exposure of patient-sensitive data, and planning to relay vulnerabilities. The reduction of risk to medical-device implants is achieved by cyberhygiene mechanisms, which involves industry propositions on safety-by-design and by protecting diagnostic imaging systems and life-intelligent network and monitoring systems and networks against malware.

5.1. Water infrastructure security

Tangible threats to the water security of nations range from man-made problems, such as asset integrity, SCADA/IoT exposure, insider action, and cyber-attacks, to natural disasters. These concerns highlight the need for a proactive stance to deal with water-buffering capacity loss, facility performance degradation, resource scarcity, and eventual supply failure. In the face of growing crisis management, timely solutions are of paramount importance. Recent incidents and patterns reveal that existing Security Contingency Plans do not adequately address the above and warn of inadequacies in IT/OT network monitoring.

There is no common baseline for the security of water pipelines and their automated control systems. The Cantor Digital-Twin-Based Architecture addresses this challenge and proposes a framework involving a cyber-attack preparedness Use Case linked to the architecture. remote controlled hindrances in several business systems merge with SCADA systems. They facilitate real time operations in monitoring, control

and managing assets [19,21-22]. There is low awareness of the risky SCADA/IoT exposure, whereas the active part of the global water infrastructure represents approximately 95 percent of the SCADA/IoT infrastructure, and it is sensitive to cyber exposure as well.

5.2. Energy system cybersecurity

Multiple recent incidents involving the hacking of cyber defenses of the energy system across the world suggests the dire need to improve the resiliency of information and communication technology (ICT) systems upon which the work of the critical infrastructure is based. While a coordinated cyberattack on the power grid has attracted widespread attention in recent years, there are growing concerns that ransomware campaigns targeting energy companies may have knock-on supply-chain effects that disrupt the entire set of infrastructure services needed to support civilian society. A Digital-Twin based capability is therefore required to harden energy networks against the threat of cyberattacks.

All these impacts on network availability and resilience can be exacerbated if preventative action is not taken to mitigate the threats. Consequently, the Digital-Twin cyber-resilience strategy proposed in Section 4.3 is framed in the context of energy networks to highlight specific areas where more intensive security-hardening work is necessary. The energy network exposure to cyber risk should ideally be reduced across the entire infrastructure. The focus should be then provided on limiting the area of attack on OT systems coupled with building strong detect-and-respond services in case of successful breaches. Also, any possibility of collateral damages to ICT or even SCADA systems should be reduced to reduce threat spread over the whole of the critical infrastructure.

5.3. Healthcare network and device security

Cybersecurity breaches have exposed sensitive medical data and disrupted clinic operation with major impact on patient health as well as the financial performance of healthcare providers. Cyberattacks on the health sector increased by 95% in 2021 and targeting of connected medical devices will continue to accelerate; both are related to improved efficiency, cost reduction, and resource challenges. Hospitals, surgery centres, and clinics are all under scrutiny. Security flaws in medical devices may enable remote control of the devices and jeopardize the privacy and physical safety of patients—potentially with lethal consequences.

Patching of vulnerabilities can take a long time, if it occurs at all. A comprehensive digital-twin approach to healthcare cybersecurity embraces clinical, medical-device, and patient-flow considerations. Telemetry and ML are applied to monitor both the network perimeter and the security of medical devices. Simulating patient flow and medical-device occupancy enables risk exposure assessment and supports decision making for resource allocation in real time [11,23-25]. An early-warning system for cybersecurity threats leverages telemetry information across the healthcare networks, and the overall healthcare governance challenges associated with this converged network security strategy receive due attention.

5.4. Transportation and logistics cybersecurity

Transportation security concerns in the ICTs, networks, and interfaces that constitute transportation security are concluded. The impacts on the related freight and supply-chain security are also described. To address potential cascating, life-threatening failures, the integrated implementation of cybersecurity is suggested in the design and management of such assets.

Information and telecommunication (ICT) systems, networks, and interfaces form an integral part of rail transport operations in many countries. Functionality traditionally supported by separate and largely isolated communications and control systems is increasingly coordinated across a common infrastructure. It also has remote condition-monitoring sensors, data-based predictive maintenance, internet protocol on-board system and dynamically customer-facing applications that also connects the freight and passenger systems to enhance efficiency [26-28]. Traffic management systems are now becoming interconnected with each other and jurisdiction to jurisdiction so that situational awareness and response can occur across borders. Road transport is becoming highly interconnected between systems, vehicles and devices connected it to the internet, where the implementation of vehicle-to-vehicle, vehicleto-infrastructure, and vehicle-to-cloud systems is expected.

According to the research, interconnectedness, interoperability, and dependencies of transport ICT systems and networks have predisposed them to external risk and the observed possibilities of cascading failures between rail networks and interfaces. Effective measures must therefore be implemented to manage cyber risk and cyberattacks threatening the safe and reliable operation of these systems. Security deficiencies in intelligent transport systems may also affect vehicle-to-vehicle and vehicle-to-smart infrastructure communications, jeopardizing transport safety. The considerations are also applicable in other permanent transportation systems like aviation and maritime whereby, safety-critical systems can fail due to cyber-incidents or because of the cascaded effects through the freight supply chain..

6. Governance, ethics, and risk management

Data governance issues have an important place in considering digital management of critical infrastructure. Digital twins which model two or more sources of information need quality provenance data which can be traced using integration and fusion pipelines to make reliable decisions. There must be data quality assurance, access control, audit, and stewardship that will be required to ensure the trustworthiness and the functions will have an opportunity to change with the maturity of the digital twin. Apart from legal requirements to comply with data privacy regulations like the European Union's General Data Protection Regulation, privacy-enhancing technologies allowing safe but useful exchange of health data with a minimum of disclosure risk are also important for medical environments in order to promote AI/ML developments. In addition, the safety of critical infrastructure cannot be neglected and must be an on-going consideration, complementing and extending availability requirements. Finally, ethical considerations pose additional challenges that must be dealt with.

Risk identification and mitigation processes may be adapted from established processes such as ISO 31010 and ISO 31000. Hazard analysis needs to include emerging threats, and the effectiveness of key controls should be validated by red teaming simulations, penetration tests, and threat modelling. Residual risk must also be managed: current approaches by combining digital tools with crisis management and incident response support appear promising and deserve further attention.

6.1. Data governance for digital twins

A digital twin requires digital-data provenance to ensure data quality and authenticity that may affect the correctness and reliability of the digital twin. The integrity of the digital twin need to be established particularly when modelling and propagating risks across digital twins for the prediction of a future-stressed events. The provenance management framework needs to be considered when ingesting data to facilitate the capability to answer who created the data, when was it created, where was the data created,

how was the data produced and why was the data created. Due to the importance of having a digital twin, researchers have also assessed samplers and online calibration in practical air pollution modeling applications. Digital twins require real-time data to keep the model accurate. However, for the direct modeling of correlations, either other model outputs or part of measurements must be used. Consequently SAMs principally tackle the calibration issues for these models.

The data for a digital twin must be complete, accurate, reliable, up-to-date and free of any bias. A generic index to describe the data quality can be defined based on the different database elements. Data access control schemes can be built by considering the owners of the datasets and the roles of the different stakeholders. It is also important for the digital data provided to be complete, since missing values are common in databases. It is crucial to control data for anomalies. Anomalies generally remain in datasets even after scrubbing. As anomalies inevitably remain in datasets, anomaly detection is often applied to produce better analysis from dirty data. Quality control of datasets, hard-code rules on important data fields, removing outliers, and cross validation are employed to reduce the anomaly problem. Good data stewards and audit maintain data quality and consistency.

6.2. Privacy, safety, and compliance

Digital twins involve decision making by use of real-world data which can be sensitive or personally identifying. This is because anonymizing datasets reduces the risk of privacy by eliminating or masking identifiers, but inappropriate methods may result in re-identification. The identification of individuals is not possible by models that learn when the data is not identifiable but special methods are required when the data is linked with publicly known properties. In the case of healthcare application, involving the use of digital twins to model healthcare-service processes and devices, the privacy of data also requires the anonymity of patient-data provenance. The issue of privacy-control should be considered as a way of balancing the level of anonymization, the performance of an algorithm as well as the intent of the disclosure.

Cybersecurity frameworks can generally identify data-safety issues and required preparations, such as threat modeling, risk assessment and management, and incident-response planning. However, revealing the security architecture underlying dedicated systems increases safety concerns. The health infrastructure is ensured by device-related regulations such as the FDA pre-market approval of medical devices facilitated by embedded digital twins and the security agenda of the Health Insurance Portability and Accountability Act in the USA [29-32]. The studies propose empowering technologies on certifying digital twin in areas such as healthcare, aviation, and Internet of Things.

6.3. Risk assessment and mitigation strategies

For all the inherent hazards that can surface throughout the lifecycle of modern society's critical infrastructures—both natural and anthropogenic—the risks associated with the anomalous effects of cyberspace on the perturbation of the operability or abandonment of one or several critical infrastructures (CIs) remain, arguably, the least known. Although considerable research effort has been devoted to assessing and quantifying the risk of damage to CI assets, including the consequences of the possible introduction of malicious, supply-chain-based attacks, a clearer understanding of the risk management strategy that is most effective in ensuring the protection of the entire infrastructure system has yet to take shape. Risk management strategies can be classified as three-level hierarchical control systems based on hazard identification and analysis, control validation, and a residual risk methodology. The

implementation of multi-level dynamic control principles—Hierarchical Control Management (HCM) Principles—should warrant the design of HCM strategies that can guarantee adequate performance in the aftermath of internal disruptions and the unintentional seizure of a danger stage.

Advances achieved within the area of cybersecurity are providing the capability for the development of intelligent cyber solutions that can conceptually establish a parallel methodological path to guarantee a Home Land Security (HLS) approach to the entire range of dangers that can affect portable critical infrastructures. The focus has therefore been redirected toward the adoption of Risk Identification-Attack Strategies for Disruption Control (RIAS-DC) solutions that can assess how damage to the cyber assets of an infrastructure risk can be transitioned in a controlled manner into a damage scenario with the reduced residual risk.

7. Challenges, limitations, and future directions

Within the last ten years, the concept of digital-twin has become common in education, business, and the services industry and enjoys the advantages of the increased usage of the Internet of Things and cloud computing. The creation of digital twins offers interesting prospects in improving the security of critical infrastructure, but there are still some major challenges. They are due to the challenges in technology created around the obstacle to formation and integration of digital twins, as well as the necessity to add predictive-analytics systems that can transmit real-time alerts and decision-support directions. There are a number of restrictions unique to cybersecurity that are also worth taking note[31,33,34]. Most of the existing real-time AI platforms do not have high levels of justification using live data and efficiency of monitoring platforms that utilize threat intelligence feeds and alarm-management systems might be incident-specific and situation-driven.

The areas of water, energy, healthcare, and transport infrastructure feature prominently, both because much progress has already been made and because additional research and development will garner high rewards. Many other physical and digital development infrastructures also merit attention. Advances in the digital-twin concept should underpin the cybersecurity of smart campuses, tourism destinations, cities, and regions, as well as of smart industries, energy, and water resilience strategies, and smart society models. Engaging in work on resilience also needs to be in the future, such as the representation of alternative scenarios related to threats by digital twins, testing, and validation of them or running stress tests in the real-time that activate parallel alert-generation procedures and dynamic status feedback of real-time AI systems.

8. Conclusion

There is significant potential of digital twins in supporting predictive analytics, early-warning systems, and real-time AI monitoring to facilitate cybersecurity. A combination of these technologies may help to secure specialized digital-asset networks which belong to vital infrastructure, such as water supply, energy, healthcare, and transportation systems. These systems should be in form of defence in depth, and layered (with several overlapping layers) and various control mechanisms like physical, technical, and human.

Digital twins of critical infrastructure require trustworthy data to assist in the early warning and identification of anomalies. Architectures for hazard groups integrity in real-time are to be detailed together with requirements of built in threat-intelligence wheel feeds. The threat of security incidents and the possibility of one certifying the resilience of the warning-system to the attempts of exploiting it

require distinguished measures. The provenance of data and its quality, the right to access, and the derogations have to be controlled with the help of the complex framework that guarantees the whole digital-twin ecosystem remained confidential and regular.

References

[1] Park W, Kwon H. Implementing artificial intelligence education for middle school technology education in Republic of Korea. International journal of technology and design education. 2024 Mar;34(1):109-35.

[2] Tedre M, Toivonen T, Kahila J, Vartiainen H, Valtonen T, Jormanainen I, Pears A. Teaching machine learning in K–12 classroom: Pedagogical and technological trajectories for artificial intelligence education. IEEE access. 2021 Jul 19;9:110558-72.

[3] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.

[4] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.

[5] Polak S, Schiavo G, Zancanaro M. Teachers' perspective on artificial intelligence education: An initial investigation. InCHI conference on human factors in computing systems extended abstracts 2022 Apr 27 (pp. 1-7).

[6] Lampropoulos G. Combining artificial intelligence with augmented reality and virtual reality in education: Current trends and future perspectives. Multimodal Technologies and Interaction. 2025 Jan 28;9(2):11.

[7] Mumtaz S, Carmichael J, Weiss M, Nimon-Peters A. Ethical use of artificial intelligence based tools in higher education: are future business leaders ready?. Education and Information Technologies. 2025 Apr;30(6):7293-319.

[8] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.

[9] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.

[10] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.

[11] Feigerlova E, Hani H, Hothersall-Davies E. A systematic review of the impact of artificial intelligence on educational outcomes in health professions education. BMC Medical Education. 2025 Jan 27;25(1):129.

[12] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.

[13] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.

[14] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..

[15] McDonald N, Johri A, Ali A, Collier AH. Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. Computers in Human Behavior: Artificial Humans. 2025 Mar 1;3:100121.

[16] Topali P, Ortega-Arranz A, Rodríguez-Triana MJ, Er E, Khalil M, Akçapınar G. Designing human-centered learning analytics and artificial intelligence in education solutions: a systematic literature review. Behaviour & Information Technology. 2025 Mar 16;44(5):1071-98.

[17] Huang X. Aims for cultivating students' key competencies based on artificial intelligence education in China. Education and Information Technologies. 2021 Sep;26(5):5127-47.

[18] Lampou R. The integration of artificial intelligence in education: Opportunities and challenges. Review of Artificial Intelligence in Education. 2023 Aug 18;4:e15-.

[19] Pham ST, Sampson PM. The development of artificial intelligence in education: A review in context. Journal of Computer Assisted Learning. 2022 Oct;38(5):1408-21.
[20] Chen X, Zou D, Xie H, Cheng G, Liu C. Two decades of artificial intelligence in education. Educational Technology & Society. 2022 Jan 1;25(1):28-47.
[21] Ouyang F, Jiao P. Artificial intelligence in education: The three paradigms. Computers and Education: Artificial Intelligence. 2021 Jan 1;2:100020.
[22] Lameras P, Arnab S. Power to the teachers: an exploratory review on artificial intelligence in education. Information. 2021 Dec 29;13(1):14.
[23] Zhai X, Chu X, Chai CS, Jong MS, Istenic A, Spector M, Liu JB, Yuan J, Li Y. A Review of Artificial Intelligence (AI) in Education from 2010 to 2020. Complexity. 2021;2021(1):8812542.
[24] Stolpe K, Hallström J. Artificial intelligence literacy for technology education. Computers and Education Open. 2024 Jun 1;6:100159.
[25] Huang L. Ethics of artificial intelligence in education: Student privacy and data protection. Science Insights Education Frontiers. 2023 Jun 30;16(2):2577-87.
[26] Kamalov F, Santandreu Calonge D, Gurrib I. New era of artificial intelligence in education: Towards a sustainable multifaceted revolution. Sustainability. 2023 Aug 16;15(16):12451.
[27] Huang J, Saleh S, Liu Y. A review on artificial intelligence in education. Academic Journal of Interdisciplinary Studies. 2021 May;10(3).
[28] Su J, Ng DT, Chu SK. Artificial intelligence (AI) literacy in early childhood education: The challenges and opportunities. Computers and Education: Artificial Intelligence. 2023 Jan 1;4:100124.
[29] Idroes GM, Noviandy TR, Maulana A, Irvanizam I, Jalil Z, Lensoni L, Lala A, Abas AH, Tallei TE, Idroes R. Student perspectives on the role of artificial intelligence in education: A survey-based analysis. Journal of Educational Management and Learning. 2023 Jul 24;1(1):8-15.
[30] Tapalova O, Zhiyenbayeva N. Artificial intelligence in education: AIEd for personalised learning pathways. Electronic Journal of e-Learning. 2022;20(5):639-53.
[31] Alam A. Possibilities and apprehensions in the landscape of artificial intelligence in education. In2021 International conference on computational intelligence and computing applications (ICCICA) 2021 Nov 26 (pp. 1-8). IEEE.
[32] Limna P, Jakwatanatham S, Siripipattanakul S, Kaewpuang P, Sriboonruang P. A review of artificial intelligence (AI) in education during the digital era. Advance Knowledge for Executives. 2022 Jul;1(1):1-9.
[33] Doroudi S. The intertwined histories of artificial intelligence and education. International Journal of Artificial Intelligence in Education. 2023 Dec;33(4):885-928.
[34] Nguyen A, Ngo HN, Hong Y, Dang B, Nguyen BP. Ethical principles for artificial intelligence in education. Education and information technologies. 2023 Apr;28(4):4221-41.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 4: AI as a Weapon: The Dark Potential of Artificial Intelligence in Cyber Threats

1. Introduction

Artificial intelligence (AI) is considered a weapon when applied to a cyber threat that uses AI in some form to increase its effectiveness compared to a similar threat that does not. Such use of AI is distinctive due to the fact that AI technologies can support four foundational capabilities: autonomy, adaptability, data-rationality. The interplay among these four capabilities can lead to a significant increase in the attack surface of society as a whole, as the attackers become outsourcing parts of the threat against their enemies. Any use of AI supporting these capabilities makes a cyber threat as strong as its use in a deterrent capability of the threat actor and, at the end of the day, can shift the overall balance among oppressive regimes and liberal democracies.

The interest of nation-states in AI is shifting beyond using it as a supporting tool towards AI as a competitive weapon in the cyberwarfare space, blurring the boundary in the actor of the cyberoperations. Nation-state actors and other large organizations already adopted AI as a supporting tool in designing and implementing large and complex cyberattacks, as their purpose is to maximize resources and minimize costs. So-called house-based agents and hybrid bionic defense strategies are already in use as defensive measures for supervision and protection AI2. In this asymmetric game, the bad actors crafty and design AI workers able to manage the SMEs' economic backdoor for ruining the competition. AI, as final player in the cyberwarfare and ready for action, is already presented in the literature for some main cyber-threats evolution directions: malware, social engineering, botnets development, ransomware; actors capable of employing these cyber tools for undermining liberal and democratic regimes for the AI-in-the-loop-attack model.

2. Foundations: AI Capabilities and Cyber Threat Landscape

AI can operate as a weapon like any other traditional tool. A weapon, or tool of destruction, is something that is intended to be used and programmed to kill, frequently for the aim of destroying or implanting oneself among other weapon-like beings. The defining characteristic of a weapon is its ability to act independently in directing its controlling user. Even if a weapon may be able to act independently on its target or act as a lackey while increasing the weapon's ability to function in the absence of specific capabilities, the weapon's goal is what distinguishes it from other tools capable of perpetrating cyber threats [1-3]. These characteristics are still especially important in the cyber realm, even though cyber threats are not enabling the destruction of hack-like viruses made in nature, weapons or tools of destruction are still available and specifically shaped.

Artificial Intelligence (AI) is one of the technologies that could have such capabilities when applied properly. But AI is not the sole technology that can accomplish this; recent developments in Generative Artificial Intelligence (GenAI) and significant breakthroughs in other key technologies have made the

nature of cyber threats much more complex than a few years ago. For instance, the number of recorded data breaches has been rising year after year, reaching a staggering peak of around 50 billion records in 2021. GenAI-based social engineering attacks made using ChatGPT have been widely reported. The hacked data of the nearly 200 million users of the popular video application TikTok and the release of voice-based automated phishing tools by a research firm clearly indicate that these types of threats are still real. AI can ultimately enable the creation of weapons for cyber viruses like the cyber weapon creator tool released in the game Watch Dogs Legion.

3. AI-Generated Malware and Self-Evolving Threats

The ability of artificial intelligence to generate malware and self-evolve adds to the existing dynamics of cyber conflicts. Both autonomous cyber operations and AI assistance in developing, deploying, or operating malicious code will be examined.

Malware capable of designing and producing new malware for data infiltration is currently evident, although it still requires human input on active defense. Self-modification and self-Persistence of malicious code have been documented as well. An independent Autonomy Loop, involving feedback from the assigned environment, and subject to reinforcement learning dynamics, can be implemented in malware operations. Such conditions enable the development of similar but distinct objects to the same task or of different tasks to the similar object, and varying operational space but approaching the more adequate data to the environment [2,4,5]. These new advances are distinguished by two distinct features: the construction of malware with the help of the existing code in the target machine and autonomy in self-updates without the further interaction of the user operator.

The capacity of code-generation, self-design, and self-modification has been reported. The common tricks of making malware more challenging to detect or to depress A/V programs have been enumerated and effectively tested. Payloads may also undergo self-modification based on an environment of choice (e.g. in virtual environments, decryption of code may not be required anymore). Self-Persistence and–Determination functions, granting the capability of reproducing a persistent copy on the same domain or changing domain when deleted, have been employed. It is also possible for payloads to include techniques for autonomously receiving additional functions for new malicious purposes. All these features are present in many deployed families, such as the known Astaroth and Chacha malware no evolution dynamics or independence from the human user/operator are still required.

4. AI-Assisted Social Engineering

The social engineering approach is improved through the use of AI to augment conventional methods with scalable outreach. Artificial intelligent is able to produce a text, voice, and video recordings. These abilities are relevant to carry out phishing campaigns, create realistic identities, initiate targeted attacks of deception, and contact people on a large scale through automation, which will provide a better chance of succeeding. For instance, an attacker can use AI-generated emails to reach hundreds of thousands of people and easily create synthetic personas to fool the few who show interest.

Brand awareness and user training programs do not help stop phishing since it remains one of the most effective attack vectors, mainly due to its ability to take advantage of human cognitive biases. Given that AI reduces the price of high-quality fabrication and allows acting at scale so that the cognitive biases it exploits are considerable, the two important questions are: Which cognitive biases do AI-driven phishing use? and What makes users victims of thrilling synthetic communication? The study of the impact of

various language styles in phishing may give information on the vulnerability as well as the literature of the possibility to identify the AI-generated fake news in social network can also bring information on the susceptibility of phishing but has not been studied yet.

The computers have become effective in producing computer-generated text responses that are human like. An example may be the ChatGPT that can converse in natural language on virtually any subject at nearly the understanding level of humans [6-8]. Voice synthesis capabilities have also improved fast with applications such as Descript having the ability of cloning the voice of users and intonation with just a few seconds of training information. Advanced art is voice cloning videos of candidates spreading fake information- effectively, deep voices..” Cloning users’ faces with just 10 images is now commercially available and enables realistic videos. The evolution in the accessibility of AI for voice and video synthesis, coupled with the fact that interest in phishing continues to morph into more efficient forms of impression management, has ushered in greater use of this deceitful method. The moment that highly convincing synthetic channels are made more easy to manufacture users will be more enticed to pick based on convenience as opposed to hard work.

5. AI-Driven Botnets and Algorithmic Ransomware

Ransomware is a computer virus that blocks access to data or computer until one pays a ransom. Although numerous ransomware variants base their work on human strategy and operational abilities, various CaaS ransomware families provide the implementation of individual payloads that are distributed and replicated in other campaigns. Operation support is then taken for aid in installation and execution by tapping external resources.. However, these actors appear to lack basic threat-awareness sophistication and tailored environments. AI more broadly suggests the potential for even greater levels of automation, not only in payload construction but also in decision-making related to actor control, resource allocation, and supply-side responses. This incorporation of AI should extend to the processes of target selection and threat deployment within the overall market and facilitate service delivery through automation at scale. External AI offer a command-and-control capability to integrate outstanding subsystems, prompt resource reclamation across the botnet, and stealth during operation. This command-and-control activity is combined with a propensity for enhancement at a system level – enabling capacity and stealth to be added during lifetime from external agents capable of boosting scalability, system checks, C2 camouflage, and brand identity modification – to enable adaptation and emulation capability on demand.

Ransomware families such as XXXL then take this AI-driven model a step further by automating creation and configuration of dedicated and tailored payloads for prospective buyers. Incremental changes that can enhance demand-side supply are similarly identified and addressed, and the CaaS delivery model is extended by opening an API marketplace for additional emulated supply-side services. Such capacity for innovation further increases the risk of catastrophic consequences from hostile action across this market, with little information suggesting these measures have been undertaken, much less been successful. Ransomware-as-a-service probably remains limited to human actors, but evidence points toward an incipient trend – with Hostivent, SmartBunker, and others demonstrating heightened automation through the generation of different kinds of traditional malware. Sprawl in these automated ransomware families is already evident, with an increasing number of publicly available automated ransomware variants that can be invoked by operators of little-to-no skill – just as is seen in the traditional CaaS ecosystem for crimeware payloads [9,10].

6. State-Sponsored Actors and Strategic Implications

Like other technologies, Artificial Intelligence (AI) has dual-use characteristics. Proponents see increasing military investment in AI, especially by major global powers, as a potential strategic revolution in warfare. AI democratizes access to advanced military capability. One could expect state-sponsored actors to apply this dual-use technology for nation-state interests. A gap in the preparation, strategy, and governance of cyber capabilities on the part of governments, businesses, and society appears to allow a first-mover advantage to threat actors [11-13]. It is a matter of time before nation-state interests weaponize AI, including advanced capabilities such as automated decision-making, decision-support systems, authentic imitation of human and machine interactions, and self-evolving capabilities. For these reasons, nation-state interests have been selected as the focus.

To determine the most dire and immediate effects of the AI technologies employed to cyber capabilities, it is informative to take into consideration the profile of threat posed by each of the actors. The motives (intentiments), capabilities (means) and thresholds of weaponization of an actor offer good analysis tool. Significant motives can include economic, power, control, ideology, and revenge, but the current focus is on the economy, ideology, and revenge. AI technologies can strongly increase capability. The state-sponsored use of AI for cyber-espionage with the aim of collecting commercial information is probably the most advanced use case for AI, where the threat from corporate espionage is, therefore, also the highest. AI for facilitating indirect revenge cyber-attacks is also at a mature stage of development, possibly through support from releases of AI malware and AI reinforcement-learning taskbots. Industry also helps keep revenge threats active through the sloppy handling of stolen private information.

The AI-enabled cyber kill chain narrows thresholds. The cost of scaling automated cyber-attacks is diminished [2,14-17]. The current military use of AI-based decision-support systems suggests that the next threshold for state-sponsored actors is advanced decision-support systems for supporting attacks in sensor/actuator-rich environmental contexts. The rapidly increasing interest of less developed countries in launching disruptive technology with AI might suggest a willingness to assume risk.

7. Defense and Mitigation: Governance, Policy, and Technical Controls

The governments should enhance the accountability of the AI associated security risks with governance systems, e.g., the Europeans commission AI act along with the proposals that identified risk categories in regulated use-cases. At the individual level, the level of current and future AI-related risk shall be factored by the organizations in pre-existing practice of risk-management process, which are required by law, regulation, or best-practice adherence and implement better and more recent detection, response, attribution, and recovery approaches with respect to people, processes, and technology [9,18-21]. Risk assessments with exposure to AI risks should be conducted using a multi-tiered approach and these should consider actor capabilities and motivations, the presence of historical indicators of compromise and incident narratives, and test whether sufficient defenses exist for methods seen in the wild and for AI-assisted attacks.

Reporting, avoiding and counteracting the abuse of AI will strain even overloaded cybersecurity centers. One should focus on the controls and mitigations that can assist in reducing assignment and detection challenges. Such controls include a zero trust architecture that helps avoid the unauthorized harvesting of extensive user profiles for initiating synthetic communications with AI-supported humanlike deception; counter-narrative programs that disrupt perennial phishing campaigns staged by actors whose motivation is the exploitation of cognitive bias rather than policy-breaking temptation; the adoption of attribution

frameworks that enable a better understanding of why, when, and how nation-states will weaponize cyberspace; and the sharing of incident data that helps improve readiness for future attacks [22-26].

8. Ethical Considerations and Risk Assessment

The analysis of the ethical factors, more particularly the dual-use, proportionality, and the acceptable risk levels, discloses the conflict between the social good and the dangerous applications of AI. To create a niche between innovative freedom and controllability, it is necessary to clarify the circumstances under which the AI technology should be designed and used to either positive or oppositional goals, according to the compatible risk-assessment strategies, and inject the technical solutions into these high principles. Such schemata demand a healthiness, expansion, and unification of structured and insightful risk-assessment systems that might be capable of reverberating the pertinent modulations, and cleverly setting aside the generation of authentic human well-being[27,28]..

The dual-use nature of AI technology and its underlying source code and algorithms create substantial conflict between societal demand for innovative progress and the related hazards. Military and paramilitary research and development occupy relatively special positions in this regard because the developing market's offer for such technology is extremely small and hardly generates the needed research dynamics. In other contexts, however, the natural tendency of markets to pursue growth manifests more easily and vigorously [29-33]. At least in this case, controlling hazardous applications while leaving intact the socially useful side of innovation requires some information processing and synthesis at the commanding levels. Proportionality and acceptable risk thresholds are practicable concepts that yield information justifying action or inaction.

9. Case Studies and Incident Analyses

Cyber operations associated with AI may not yet be observable, as the technology is neither sufficiently developed nor deployed among actors capable of such operations. However, the rapid pace of development suggests that it is only a matter of time. Some of the attacks are characterized by marks of involvement of AI albeit none have been proved to have been caused by AI so far. The study of previous cases as well as the patterns and developments may assist in determining how the technology may be used maliciously and what organization may be doing with it first.

The most probable forms of AI-based attacks in the nearest future would be the AI-enabled botnets and AI-assisted social engineering. The AI to manage the command-and-control servers and keep track of the status of all nodes in the network, congestion, probability of being caught, execution of a deployed payload, and dynamically generate the code to support the payload may orchestrate botnets [34-36]. The use of generative AI technology to generate phishing messages, voice samples, as well as facial images to impersonate the target, and automation to run the campaigns on a large-scale level have the potential to become common through AI-assisted social engineering (with common good intentions). Phishing attacks using this type of AI-only generated messages can be more effective than standard attacks because of the lack of language artifacts which frequently identify machine generated messaging, and also the weak heuristics that are typically employed to counteract the attacks.

10. Conclusion

Thesis and Analysis Structure

AI has been a game-changing concept since its inception. Recent popular application and additional popularization by the media has significantly elevated the hype construction in the modern world, and once again, the discussions about the warning issued by Einstein begin: will it be the salvation or the weapon of destruction? Demigod217 provided the first analysis of potential exploits of AI-enabled analysis. It sought to identify areas of AI progression, assess the responsible use of AI, and identify malicious use within the cyber threat framework.

All technology can be used for good and evil purposes. It is thus not too difficult to envisage the weaponization of AI in ever-increasing malicious applications. This paper confronted the hypothesis that AI is a weapon of destruction in the cyber-threat spectrum, thus offering an overview of foundations and prime contenders for AI-enabled cyber threats as they currently stand. Relevant existing capabilities were identified alongside their limitations and the current level of enabling technologies required for malicious deployment. Using those foundations as a base, advancement indicators for the generation of AI-enabled offensive threats were then described through a context-based analysis of four specific AI application areas: AI-generated malware and self-evolving threats; AI-assisted social engineering; and AI-driven botnets and algorithmic ransomware.

References

[1] Paek S, Kim N. Analysis of worldwide research trends on the impact of artificial intelligence in education. Sustainability. 2021 Jul 16;13(14):7941.

[2] Chiu TK, Xia Q, Zhou X, Chai CS, Cheng M. Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. Computers and Education: Artificial Intelligence. 2023 Jan 1;4:100118.

[3] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.

[4] Mohapatra PS. Artificial intelligence and machine learning for test engineers: concepts in software quality assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:17.

[5] Akgun S, Greenhow C. Artificial intelligence in education: Addressing ethical challenges in K-12 settings. AI and Ethics. 2022 Aug;2(3):431-40.

[6] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.

[7] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.

[8] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.

[9] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..

[10] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.

[11] Pratama MP, Sampelolo R, Lura H. Revolutionizing education: harnessing the power of artificial intelligence for personalized learning. Klasikal: Journal of education, language teaching and science. 2023 Aug 10;5(2):350-7.
[12] Tan X, Cheng G, Ling MH. Artificial intelligence in teaching and teacher professional development: A systematic review. Computers and Education: Artificial Intelligence. 2025 Jun 1;8:100355.
[13] Abbasi BN, Wu Y, Luo Z. Exploring the impact of artificial intelligence on curriculum development in global higher education institutions. Education and Information Technologies. 2025 Jan;30(1):547-81.
[14] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[15] **Mohapatra** PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[16] **Panda** S. Scalable Artificial Intelligence Systems: Cloud-Native, Edge-AI, MLOps, and Governance for Real-World Deployment. Deep Science Publishing; 2025 Jul 28.
[17] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.
[18] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.
[19] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science
[20] Baig MI, Yadegaridehkordi E. Factors influencing academic staff satisfaction and continuous usage of generative artificial intelligence (GenAI) in higher education. International Journal of Educational Technology in Higher Education. 2025 Feb 3;22(1):5.
[21] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[22] Yang Z, Wu JG, Xie H. Taming Frankenstein's monster: Ethical considerations relating to generative artificial intelligence in education. Asia Pacific Journal of Education. 2025 Aug 8;45(4):1330-43.
[23] Abulibdeh A, Zaidan E, Abulibdeh R. Navigating the confluence of artificial intelligence and education for sustainable development in the era of industry 4.0: Challenges, opportunities, and ethical dimensions. Journal of Cleaner Production. 2024 Jan 15;437:140527.
[24] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[25] Wang S, Wang F, Zhu Z, Wang J, Tran T, Du Z. Artificial intelligence in education: A systematic literature review. Expert Systems with Applications. 2024 Oct 15;252:124167.
[26] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[27] Mao J, Chen B, Liu JC. Generative artificial intelligence in education and its implications for assessment. TechTrends. 2024 Jan;68(1):58-66.
[28] Vieriu AM, Petrea G. The impact of artificial intelligence (AI) on students' academic development. Education Sciences. 2025 Mar 11;15(3):343.
[29] Topaz M, Peltonen LM, Michalowski M, Stiglic G, Ronquillo C, Pruinelli L, Song J, O'connor S, Miyagawa S, Fukahori H. The ChatGPT effect: nursing education and generative artificial intelligence. Journal of Nursing Education. 2025 Jun 1;64(6):e40-3.
[30] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.

[31] Shahzad MF, Xu S, Asif M. Factors affecting generative artificial intelligence, such as ChatGPT, use in higher education: An application of technology acceptance model. British Educational Research Journal. 2025 Apr;51(2):489-513.
[32] Bewersdorff A, Hartmann C, Hornberger M, Seßler K, Bannert M, Kasneci E, Kasneci G, Zhai X, Nerdel C. Taking the next step with generative artificial intelligence: The transformative role of multimodal large language models in science education. Learning and Individual Differences. 2025 Feb 1;118:102601.
[33] Malik AR, Pratiwi Y, Andajani K, Numertayasa IW, Suharti S, Darwis A. Exploring artificial intelligence in academic essay: higher education student's perspective. International Journal of Educational Research Open. 2023 Dec 1;5:100296.
[34] Baidoo-Anu D, Ansah LO. Education in the era of generative artificial intelligence (AI): Understanding the potential benefits of ChatGPT in promoting teaching and learning. Journal of AI. 2023 Dec 31;7(1):52-62.
[35] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.
[36] Adams C, Pente P, Lemermeyer G, Rockwell G. Ethical principles for artificial intelligence in K-12 education. Comput. Educ. Artif. Intell.. 2023 Apr;4:100131.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 5: AI and the Weaponization of Information: Deepfakes, Influence & Psychological Warfare

1. Introduction

The growth of Artificial Intellect (AI) technologies that can generate large amounts of artificial content due to bulk production has been an object of concern to the quality of the election process, political culture, and geopolitics. Given that the production and dissemination of misinformation are symbolic of Information Warfare, the potential for AI to enable new ways of generating and sharing false content, or of improving coordination and speed of elaborated misinformation, requires further study.

The assumption is that the convergence of AI and influence systems and misinformation systems could result in more true and believable fakes, the timing or responsiveness of the production process- i.e. able to use misinformation to look-a-like hoop visual performance as well as reduce the cost and barriers to coordination, and consequently increase the dissemination and effectiveness of politically motivated misinformation. In this regard the AI-generated content could be perceived as the type of content that is fully or partially created using the aid of AI technology that allows it to do the generative work or supporting the modeling and processing tasks such as a faster speed of work and being able to support multiple languages.

The use of human-in-the-loop Deep Learning technologies for the elaboration of misinformation could inject higher levels of customization and personalization into the process, yet could modify the way audiences suspect the reliability of the information, without necessarily making it more unreliable. Regardless of the development of more automation and sophistication, the technical obstacles of the production of high-volume influence content are still comparatively high. In comparison, the propensity of platforms to understand the information as fake is among the issues with a profound implication on the way people process online information.

2. AI-enabled misinformation

By using artificial intelligence, creating artificial content becomes easy and enhances the spread of the same content by using social-media networks. Big language models produce convincing text and the text-to images, text-to-video and audio-generation systems make multimedia creation more affordable. Bots lead to quick distribution and automated interaction, which results into the operations of information accessing geographical and linguistic barriers that were formerly difficult to capitalize on [1-4]. The content provided with AI assistance is more precise than the misinformation provided traditionally, more timely than the fact-checking, more scalable than the traditional journalism, and more persistent than the other online contents. In contrast to conventional forms of misinformation, which are limited by manual creation and linguistic fantasy, artificial intelligence-driven systems create everyone an information

economy opportunity, reduce the incentive to tell the truth, and invert social relationships of death funnels in information, acknowledging the difficulty of debunking a narrative once it has gone viral. Nevertheless, real-life happenings are the ground truth, and there is an indication that falsely designated synthetic content is commonly and accurately distinguished.

They affect misinformation through five factors, namely the advent of digital assistants that can generate language-based content, both negative and positive; the increased use of AI text generators in social-media settings; the falling cost of creating videos of cinema quality on personal computers; the proliferation of voice-cloning software; and the proliferation of AI image creation or manipulation tools. Identifying AI-generated misinformation and, in particular, when labeled is not to remove the focus on non-synthetic content with the purpose of faking elections or is a by-product that can be neglected as harmless. The problem with misinformation is that it is diluted during the election periods. The issue of showing the warning label to correctly labelled AI-generated deepfakes is not yet economical enough, placing the desensitizing of audiences under threat.

3. Deepfake attacks on public trust

In media-deepfakes, a user presents fake content in the form of videos or audio files where a person is made to seemingly say or do something that the person was not actually saying or doing. The manipulations can alter real media, generating seemingly realistic audiovisual material in which the persona of the original subject is superimposed onto another individual's face and body or their facial features are transformed using neural networks. Recent technologies on deepfakes, being inherently hyper-realistic in imagination, rely on algorithms in deep-learning based on the same, which is that the placement of facial features and disposition are correlated based on telegenic characteristics. Such generative models are able to gain and comprehend how his/her person speaks, recreating a distinguishing feature, voice quality, tone, pitch, cadence, and accent.. More subtle manipulations can change voice and facial expressions, eye gaze, head orientation, or even lip movement in any kind of audio or video, producing altered media or original material.

Even though the technical solutions to the problem of deepfakes have become so customary, like the Deepfake Detection Challenge or other similar projects aimed at Google Research or Facebook, the danger of undermining the reputation of the authority by using deepfakes, e.g., by creating false or inappropriate pronunciations as belonging to world leaders, remains to be in the center stage. These instances take advantage of media-deepfakes in life-or-death situations and transfer authority off the institutions and experts more rapidly than legitacy can be rebuilt. It is reported that people who believe the deepfakes more have been less disposed to doubt the reality of the material not of deepfakes, have less knowledge about the technology and experiencing less mental load at the time of exposure [5-8]. It is also hinted several times that trust in deep fake based stories is not only the cause of intent to share but also intent to share in the presence of counter-information. Bayesian updating of people’s beliefs about particular political events when they see a deepfake-skewed video increases misinformation spread and decreases issue comprehension. Deepfake presence nevertheless does not reduce the likelihood of indicating a trusting response.

4. Social-media influence operations

The influence operations of social-media are described as operations of a four-part methodological framework. The actors take advantage of the distribution platforms of the general population and focus

their attention on the user-generated content with a target on the potentially vulnerable groups. The impact content / false content, between the authentic, misleading and the truly artificial disinformation, is normally supercharged by systems of bots, pseudonymous accounts and hidden coordination. The last step is to maximize exposure through virality-seeking algorithm and amplification networks. These functions are fulfilled by both the host- and third-party services and ensure that the inauthentic behaviour is not visible.

It is difficult to gauge the effectiveness. Quite a number of operations are small and short lived, yet there are those which are rated high through the preferred metrics, and which have produced off-line impacts that are tangible. The existing information indicates trends that usually define large-scale operation but the existence of which is not a mandatory requirement of a successful attack. In their endeavors to attack such subversions, platforms are doing their best to be more asymmetric in policy, patchy in their moderation, and many platforms also have poor content-detection.

5. Defending democratic processes

Despite the fact that AI-enabled fake news boosts the performance of the crooked actors, it improves the security of democratic practice as well. The reactions of first generations to synthetic content were to alleviate the damages caused by its formal features, which are verification protocols, transparency of the source, and information literacy. The readiness activities were based on the stability of the institutions and information ecosystem [9-13]. Besides the curtailing synthetic character of electoral content, the defenses were intended to increase the reaction and response to the occurrences at other fronts. Coordination on a multi-stakeholder level, through technology, policy, media, education, and civil-society, was thus all-important as well as a strategic examination of the risk profile of AI-enabled misinformation and consequent distribution of resources. The design of disincentives should be done only at a second stage in order to address the incentives of the dishonest actors.

One of the simplest possible defenses is verification protocols. Their usefulness, however, will be related to the ubiquity, along with the functionality of the user verification as the behavior of the urgency, risk perception, and trustworthiness of other characteristics of the information source. When it comes to the sensitive situations like elections or situations of terrorism, verification might be beyond the capabilities of one single entity especially in the light of unsupportable rumors [14-18]. A system which attributes and presents an overall rating to the content of a node, using the ratings of previous consumers, would hence be useful, along with resilient two-sided markets with incentives on reliable supply of information. The attempts to avoid inauthentic conduct are also generally advantageous, but ultimately, moderation and other platform rules must be oriented to fighting the strategic power abiding infidelity, instead of the sheer existence of inauthentic accounts or the coordinated interaction patterns..

6. Ethical, legal, and policy considerations

Artificial intelligence research community has members who problematize the ethical, legal, and policy aspects of AI-enabled misinformation particularly in cases of violence and harm, freedom of speech and democracy as well as weaponization of general purpose technologies. They check background expectations, privileges and obligations and check the enforcement of the current legislative measures in creating, spreading and using contents [19-23]. Such measures as the increased deterrence, detection and redress mechanisms, the abolishment of restrictions on content moderation and upholding the unbiased platforms, and the assistance of various national and global institutional frameworks are suggested.

The stages that present some ethical, legal, and policy issues are at enabling-development, content-creation, distribution, use, and misuse levels. Some of the misuse includes violent content, intent-based violent incitement, sextortion, impersonating a brand, misinformation on elections and COVID-19, and the AI-induced self-harm, such as digital self-harm by Black adolescents targeting the beauty ideal and disinformation-based anxiety in young people. It also entails the conflict-in-loop version of misery and the toy problem of arguing with a hate-spewing chatbot, through the use of AI. General-purpose technology has been said to be at the extreme end of the scale as being those weapons which take on independent lives and the problem of free speech is a paradox where people want to be limited on mass speech in order to maintain order in the country and national security.

7. Methodologies for detection and resilience

The process of methodological development and assessment have embraced some technical methods of identifying and eliminating the occurrence and promotion of AI-based misinformation. Multimedia forensics applies the capabilities of digital-exploitation and other cross-domain detector models to identify, typically automatically, synthetically generated or edited content, typically after signals of relying on those signals. It is relevant to gain a more insightful view of the potential and constraints of such approaches in order to utilize them in a responsible manner.

It is resilience to disinformation that implies not only the technical monitoring of disinformation but a suggestion that it is important to explore tools and other resources that increase individual information literacy and institutional support systems. The political leaders, political parties, and civil society must be ready to the fact that the military operation may be suddenly enclosed in the alterations in the information environment. Any effort to come up with new resilience mechanisms should therefore be done in a systematic approach, which is methodologically rigorous. Best practices suggested to the community point at the fact that public benchmark datasets should be available so as to prove the detection algorithms introduced recently [24-28]. This is achieved through replication plans that allow independent reproduction of the already published results and achieve a more valid scientific foundation. Besides, the quantification of uncertainty must become a regular part of the new detector development, to make sure that it would integrate correctly into a larger ecosystem. The research incorporates interdisciplinary studies that combine multimedia forensics with psychometrics with political-cognition viewpoint and election application in Asia.

8. Case studies: elections and geopolitical tensions

Contemporary societies, economies and political processes are dependent on communication and information as a vital driver. They are also able to generate and coordinate chaos and manipulation on a large scale, as it happened with the recent high-stakes elections and the situation with geopolitical tension. The following section provides examples of AI-enabled misinformation - and indeed AI-enabled capability of psychological warfare, most prominently if the allowed misinformation is perpetrated and becomes apparent during the election process or during the occurrence of an international crisis - and questions who should be held responsible, who should be blamed, and who should notice when such incidents take place [29-34].

The historical and contemporary propaganda are known to be two sources of disinformation. Less forthright and mysterious, propaganda has always been an inherent aspect of war and has existed more than 1000 years. The emergence is particularly due to new technology, and AI-mediated misinformation

facilitates propaganda faster and in larger proportions than ever. The experience of these AI-supported forms of geopolitical psychological warfare will be used to work against any AI-enhanced fake news, disinformation-smear, and even confidence-manipulation.

9. Recommendations for stakeholders

Authors, platforms, journalists, educators, and technologists should all collectively protect the system of democracy against AI-related misinformation. Transparency of synthetic media in conventional and social media increases the robustness of information systems through verification guidelines on what is published and disclosed. Voter information literacy training can prevent the collective influence operation. They are empowered by active-reserve institutions, whereby they are capable of responding promptly to elections; this is the case of crisis communication teams in governments, parliaments, political parties, and military forces without this amounting to a serious political interference.

In greater scale, the active monitoring, analysis of the ecosystem and resource allocation by the collaborative networks enhance resilience further. External stakeholders can improve incident potential memetic-mapping systems for detecting and measuring activity-metric thresholds on Telegram, TikTok, and X, properly source and contextualize the activity of online-governmental institutions during electoral events, and more effectively expose and enable recovery from misuse on platforms and ecosystems not currently part of prioritized resources for securing elections. For structural mitigation, riskier times in a political cycle—such as a national election—benefit from reductions in capacity in a negative-damage mitigation approach adjusted and tuned across places and events, rather than geophysical-source-shut-down approaches [35-38].

Information sources remain a persistent area of concern. AI-assisted synthetic misinformation and influence activity are faster, more accurate, more abundant, more difficult to trigger sources and conduits, and more reliably sourced for action, adaptability, and posing than historical forms of the same. The collective harms of such messaging are rapidly amplified through the constellations of behavioral-change models applied by state-accredited social-media-advertising departments through provided access to service-supply-creation interfaces and synthetic-media-support mechanisms. Individual and collective state preparation and response for national elections, especially revolving around governmental and military-organizational agency, are important countermeasures [39-43].

10. Conclusion

The issue of resiliency in the democratic processes is still open because even the dynamics of politics and society are shifting. Verification and source-checking is essential to aspects of safeguard mechanisms to allow to find the presence of deceptive postings spread through networks of artificial accounts within a social-media platform. Equally important is the information literacy of individuals, including critical awareness and a healthy skepticism with regard to online-distributed information. However, these are insufficient alone: institutions such as media outlets, judicial systems, and democratic structures need to be robust. When a falsified audio or video is evidently manipulated and disclosed in advance, the repercussions may be limited. A primary concern of hiring agencies is how candidates respond to nondisclosure of past episodes of misconduct; even verbal accusations cloud the judgment of voters. Widespread acceptance of fabricated stimuli may in itself trigger doubts about the legitimacy of an election.

Mitigating institutional and social media vulnerabilities to this kind of AI-enabled threat requires a multi-stakeholder approach. Improvements for decision-making and allocation of available resources are feasible through ongoing evaluations of potential risks, and information warfare should be minimized before elections through better alignment of the incentives of the various stakeholders involved. The models of promoting and protecting the democratic process could be based on best- and worst-case scenarios as long as the analysis of data sets distinguishes between the real expression of the consensus and artificial influx of a message. Three ways the lawmakers and government intelligence agencies can reduce the possible risk of AI-generated misinformation affecting future political actions include H1: building proactive structures and interagency relationships to continue monitoring the digital ecosystem; H2: evaluating and investing resources to enhance the resilience of the democratic system; and H3: offering incentives to the platforms to enhance the system policies and detection solutions.

References

[1] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.

[2] Zafari M, Bazargani JS, Sadeghi-Niaraki A, Choi SM. Artificial intelligence applications in K-12 education: A systematic literature review. Ieee Access. 2022 May 30;10:61905-21.

[3] Zadorina O, Hurskaya V, Sobolyeva S, Grekova L, Vasylyuk-Zaitseva S. The role of artificial intelligence in creation of future education: Possibilities and challenges. Futurity Education. 2024 Apr 30;4(2):163-85.

[4] Wang Y, Derakhshan A, Ghiasvand F. EFL teachers' generative artificial intelligence (GenAI) literacy: A scale development and validation study. System. 2025 Jul 25:103791.

[5] Wang C, Wang H, Li Y, Dai J, Gu X, Yu T. Factors influencing university students' behavioral intention to use generative artificial intelligence: Integrating the theory of planned behavior and AI literacy. International Journal of Human–Computer Interaction. 2025 Jun 3;41(11):6649-71.

[6] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.

[7] Tuygunov N, Samaranayake L, Khurshid Z, Rewthamrongsris P, Schwendicke F, Osathanon T, Yahya NA. The transformative role of artificial intelligence in dentistry: a comprehensive overview part 2: the promise and perils, and the international dental federation communique. International Dental Journal. 2025 Feb 25.

[8] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.

[9] Su J, Yang W. Artificial intelligence in early childhood education: A scoping review. Computers and Education: Artificial Intelligence. 2022 Jan 1;3:100049.

[10] Sperling K, Stenberg CJ, McGrath C, Åkerfeldt A, Heintz F, Stenliden L. In search of artificial intelligence (AI) literacy in teacher education: A scoping review. Computers and Education Open. 2024 Jun 1;6:100169.

[11] Simms RC. Generative artificial intelligence (AI) literacy in nursing education: A crucial call to action. Nurse Education Today. 2025 Mar 1;146:106544.

[12] Shivadekar S, Mangalagiri J, Nguyen P, Chapman D, Halem M, Gite R. An intelligent parallel distributed streaming framework for near real-time science sensors and high-resolution medical images. In50th International conference on parallel processing workshop 2021 Aug 9 (pp. 1-9).

[13] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.

[14] Shidiq M. The use of artificial intelligence-based chat-gpt and its challenges for the world of education; from the viewpoint of the development of creative writing skills. InProceeding of international conference on education, society and humanity 2023 May 30 (Vol. 1, No. 1, pp. 353-357).

[15] Shata A, Hartley K. Artificial intelligence and communication technologies in academia: faculty perceptions and the adoption of generative AI. International Journal of Educational Technology in Higher Education. 2025 Mar 14;22(1):14.
[16] Sabri H, Saleh MH, Hazrati P, Merchant K, Misch J, Kumar PS, Wang HL, Barootchi S. Performance of three artificial intelligence (AI)-based large language models in standardized testing; implications for AI-assisted dental education. Journal of periodontal research. 2025 Feb;60(2):121-33.
[17] Ruiz-Rojas LI, Acosta-Vargas P, De-Moreta-Llovet J, Gonzalez-Rodriguez M. Empowering education with generative artificial intelligence tools: Approach with an instructional design matrix. Sustainability. 2023 Jul 25;15(15):11524.
[18] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.
[19] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.
[20] Rodríguez-Ruiz J, Marín-López I, Espejo-Siles R. Is artificial intelligence use related to self-control, self-esteem and self-efficacy among university students?. Education and Information Technologies. 2025 Feb;30(2):2507-24.
[21] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[22] Rahiman HU, Kodikal R. Revolutionizing education: Artificial intelligence empowered learning in higher education. Cogent Education. 2024 Dec 31;11(1):2293431.
[23] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[24] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[25] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..
[26] Owan VJ, Abang KB, Idika DO, Etta EO, Bassey BA. Exploring the potential of artificial intelligence tools in educational measurement and assessment. Eurasia journal of mathematics, science and technology education. 2023 Aug 1;19(8):em2307.
[27] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.
[28] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.
[29] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.
[30] Nemorin S, Vlachidis A, Ayerakwa HM, Andriotis P. AI hyped? A horizon scan of discourse on artificial intelligence in education (AIED) and development. Learning, Media and Technology. 2023 Jan 2;48(1):38-51.
[31] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.
[32] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[33] Lin CC, Huang AY, Lu OH. Artificial intelligence in intelligent tutoring systems toward sustainable education: a systematic review. Smart learning environments. 2023 Aug 28;10(1):41.

[34] Kurian N. AI's empathy gap: The risks of conversational Artificial Intelligence for young children's well-being and key ethical considerations for early childhood education and care. Contemporary Issues in Early Childhood. 2025 Mar;26(1):132-9.
[35] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[36] Kamila MK, Jasrotia SS. Ethical issues in the development of artificial intelligence: recognizing the risks. International Journal of Ethics and Systems. 2025 Jan 30;41(1):45-63.
[37] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.
[38] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.
[39] Hardaker G, Glenn LE. Artificial intelligence for personalized learning: a systematic literature review. The International Journal of Information and Learning Technology. 2025 Jan 13;42(1):1-4.
[40] Mohapatra PS. Artificial intelligence and machine learning for test engineers: concepts in software quality assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:17.
[41] Gligorea I, Cioca M, Oancea R, Gorski AT, Gorski H, Tudorache P. Adaptive learning using artificial intelligence in e-learning: A literature review. Education Sciences. 2023 Dec 6;13(12):1216.
[42] Gkintoni E, Antonopoulou H, Sortwell A, Halkiopoulos C. Challenging cognitive load theory: The role of educational neuroscience and artificial intelligence in redefining learning efficacy. Brain Sciences. 2025 Feb 15;15(2):203.
[43] George B, Wooden O. Managing the strategic transformation of higher education through artificial intelligence. Administrative Sciences. 2023 Aug 29;13(9):196.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 6: Artificial Intelligence in American Intelligence: GEOINT, SIGINT, and Predictive Defense

1. Introduction

Artificial intelligence (AI) is one of the tools that are integrated into the American intelligence in order to help in collecting, analyzing, and predicting. There are three main areas of application that include geospatial intelligence (GEOINT), signals intelligence (SIGINT), and predictive defense. American firms have set or proved competencies in all of these segments. These capabilities and related technical, operational, and policy issues along with the concerned terminologies are summarized in the analysis.

GEOINT capabilities that are based on artificial-intelligence are automated satellite image detection and analysis, including object detection, image segmentation, and object change detection; and pattern-of-life modeling which identifies the pattern and behavioral movement of physical objects. Although these capabilities resemble traditional modes of intelligence collection, they offer the potential for dramatic shifts in the pace and volume of warehouse collection and analysis. The dangers of increased collection and analysis rate, especially the desire to formulate and respond to conclusions instead of assumptions, are another aspect that should be taken into account concerning the given and similar abilities in other areas.

2. The Integration of Artificial Intelligence into GEOINT

There are varieties of algorithms which have been invented to computerize the analysis of satellite images. They are as simple as the detection and segmentation of objects to the more sophisticated methods of detecting subtle changes. There are object detection algorithms and image segmentation algorithms which automatically detect and define significant features of an image, and could be trained using labelled images provided by the National Geospatial-Intelligence Agency and the likes. Detecting the changes in the world is a task of the change detection algorithms that build on the developments made in the field of computer vision available to detect how the world is changing and what is being built, destroyed or transformed. The algorithms could be tested and compared to the traditional manual testing when the measures of accuracy are defined and agreed on i.e. precision, recall and intersection over union [1-4]. The validation usually takes advantage of the old tradition of scratching the surface of the services offered in this field in order to determine how the analysis can be automated and scaled.

Another research that is highly promising in the future models the dynamics of the daily life in places of interest. Movement, schedules and behavioural data can be analysed to form a model of normal life as well as identifying abnormalities of the model. These methods contrast these methods of object detection and segmentation because they are not visual models; they maintain movement using a lower-dimensional

representation. The advantage of process patterns of activity is that this enables the automated systems to identify misplaced activity: when two individuals are continuously in the same location at the same time, a continuing absence of either or both may mean something. These models should be ratified. Unlike a real-world data, they are usually tested only to demonstrate that a pattern is detected correctly in case of known a priori.

2.1. Automated Satellite Image Analysis

The automatic detection, segmentation, and classification of objects in satellite images have become possible based on machine learning techniques, especially convolutional neural networks. These capabilities are augmented by change detection algorithms which estimate material changes in a scene by use of a reference image [3-5]. Algorithms are applied to pictures that are taken by still cameras and techniques of which human analysts tend to disregard the images simply because they exist in such great numbers. Usually validation is based on some review of labeled images although as the labels are produced by analysts, they are not independent of the performance of the system being evaluated. Another scheme considers crowd-sourcing and wisdom of the crowd to produce independent visual search ground truth of small and medium-sized objects on satellite images of several commercial providers. This attempt proves that spatial scale affects the search teams of crowd workers in its accuracy. The credibility of the data, the appropriateness of the sensor in performing the acquisition tasks as well as the application of ensemble learning as a way of enhancing the classification performance are also addressed.

Although such techniques are supposed to make a big break in the field of intelligence collection or analysis, they have not done so due to the slow pace of decision making by human being that cannot fully realize the potential of automated creation of high-resolution images [6,7]. Pattern recognition heuristics tend to induce analysts to turn possible intelligence into a specific acquisition of interest instead of getting collection to certain categories of interest using pattern-of-life frameworks based on the movement, schedules, and behaviors of the real targets. Machine learning can only be fundamentally helpful in intelligence and analysis when the change-detection and tested pattern-of-life models can integrate space-based facilities with other modalities in a risk-aware manner. However, the possibility of the fundamental change in these fields is present: when there will be enough amount and quality of labeled high resolution satellite images, generative models of subject-specific, agent- and observer-centric and full-motion rendering will be developed.

2.2. Pattern-of-Life Modeling

Pattern-of-life modelling goes out into individual or group activity prediction based on the past history and the current situation based on the information, offering actionable intelligence on both national security and law enforcement effectiveness. Some of the inputs may consist of movement, work schedules, place of interest and activity profile. The mode-of-transportation prediction models based on deep learning, particularly travel-mode inference, can predict the mode of transportation i.e., walking, driving, bus, or train, with the help of trajectory data well. Examples of patterns which one can determine with the use of the temporal and spatiotemporal information include school or business hours, weekend-relaxation periods and frequencies of travelling to certain locations. Decisions that are in the short term, minutes to months ahead and pre-established areas of operation aid in formulation of the model. Such models are powerful but data hungry and thus they can perform well only when of good quality and

adequate amount [2,8-10]. It has demonstrated the possibility of real-time detection of sensitive speeches involving safety-risk traveling at a long distance.

Pattern-of-life modeling enhances the efficiency of alerts, conserves resource allocation and improves the intelligence fusion with other modalities. However, there are possibilities of making errors particularly when making predictions on non-repetitive events such as terrorist attacks. Besides, dependence on those systems and limitations placed on the action of an agent can raise the risk of intelligence failure or misinterpretation risk. Pattern-of-life modeling can be frequently tested on known cases.

2.3. Implications for Intelligence Collection and Analysis

The increasing application of AI-MLOs in the modeling of autonomously generated streams of dynamically changing geospatial intelligence of moving objects, periodically occurring rituals of activity, and dramas of complex behavior (i.e., sequences of activity rituals) can radically transform the process of analysis. In the case of long-term targets, prescriptive intelligence preparation of the battlefield (IPB) decisions can be made of G2 art and instruction and the result is one-off collection where volatile scenes are to be updated, such as a season change. Decision makers monitoring rapidly evolving events, by contrast, rely on an ever-larger “cloud of unknowing” around and above specific high-value-hardened targets whose paint-streaked sodden camouflage is no longer believed to offer meaningful cover. Against such time-sensitive dynamics, the timely digital fusion of autonomous geospatial surveillance with other modalities—SIGINT, HUMINT, and OSINT—has always been problematic, remaining a well-resourced dream or fantasy [1,11-12]. The availability of an expanding array of automated geolocation systems combining simple signal source direction-finding and time-difference position-fixing is very welcome.

Here, the pressing need is not to rely excessively on one set of AI-ML model, be it war messages on layer-2 ISP or hostile population on layer-3: figures should be seen a suspicion. Automated, agent-aided, human-in-the-loop and human-in-command methods should be taken into account and real-world interpretation should be always part and parcel of the decision-support architecture. The race to leverage terror in the form of evil motives into change-detection datasets of cognitive and off-the-shelf seasonal-activity models must be managed with discipline. As the geospatial analyst for the U.S. embassy along the brickyard-Cu Chi iron/SADAT corridor for the 1968 Tet offensive discovered to his deep regret, what can go wrong will go wrong [13-15]. And the misinterpretation of normally benign patterns of life and speech activity can have life-and-death consequences.

3. AI in SIGINT and Communications Intelligence

Communications intelligence (COMINT) and signals intelligence (SIGINT) systems powered by artificial intelligence aid human capability to analyze data by use of diverse modalities, one of them being machine learning (ML) to filter noise as well as identify new signal patterns in data streams. Moreover, non-supervised anomaly detection graphs based on labeled graph networks on traffic frequently find uses as systems that automatically classify traffic on-the-fly [16,17]. Depending on the level of high-confidence automation that can be achieved, both types of systems have the potential to enable more effective and more efficient operation and management of COMINT and SIGINT resources, including personnel.

Machine-Learning-Assisted Anomaly Detection - Machine learning within the framework of network traffic as well as voice and acoustic signature data could result in considerable operational advantages. ML-assisted anomaly detection systems and other systems that classify voiced and text messages can be evaluated based on false alarm rates, false dismissal rates, and classification accuracy for particular traffic

typesLimiting the rate of false alarming of a specific anomaly detection function not only enhances efficiency of the work force of analysts by allowing them to give more relevant alert to their attentions but also by improving overall efficiency by providing better prioritizations.

3.1. Anomaly Detection and Signal Processing

Anomaly detection Detection of patterns or events that are not congruent to a predicted behavior or distribution, is becoming a larger part of SIGINT. The necessity in high throughput in wireless communication, cyber monitoring, and sensor system support can be met using such techniques that could be supported by artificial intelligence (AI) and machine learning (ML) [12,18-20]. In wireless networks, noise reduction methods serve to increase network capacity while detection for other previously unknown signal distortion exploits per-channel SNR increase.

Anomaly detection is based on the concept of normal behavior in a system. Various detection methods outline high-dimension data representation of data streams and detection using neural networks. Methods which rely on the tatistical learning theory attempt to measure a priori relative costs of a false alarm and nondetection. Signal classification assists in the effective utilization of the available resources and adaptation to the unfavorable conditions (fog, noise and radar interference). Applications make use of preclassifiers like Neural Network, Random Forest, Support Vector Machine and Decision Tree with HMMs used to detect new classes and LDA used to learn effectively [21-23].

The performance evaluation measures, which are not recognition accuracy including precision, recall rate, and false alarm rate. Accuracy is nevertheless an important selection criterion, and also considers the recognition speed for real-time monitoring systems. Detection accuracy identifies the probability that a non-Poisson target is recognized correctly on alert.

3.2. Network Traffic Analysis and Fusion

Signals intelligence (SIGINT), especially communications intelligence (COMINT) is a domain of operation of intelligence disciplines that are concerned with collecting data through electronic communications, where telephone and internet communications constitute the bulk. AI is actively used on the collection and analysis of SIGINT, the United States is interested in creating AI technologies that will work with vast loads of heterogeneous data on communications [24,25]. The AI methods will help in different tasks, which include anomaly detection, analyzing network traffic patterns, and signal processing. Semantic anomaly detection by automatic methods on the network traffic will help to identify malware activities, DDoS attacks, and intrusions by unauthorized access into the network in a timely manner. These methods aid in reduction of the alarm response time and false alarm rates.

Communications content can be analyzed using key features obtained by automated semantic analysis and machine translation of various languages and types of communications. The development of these optimal intelligent detection methods is possible with the help of AI models that merge such features through graph neural networks and CNNs [26-28]. The models enhance proper and on time detection, and the quality lies in its varying feature fusions. Network traffic data can undoubtedly reveal transnational crime and terrorist activities, but it is frequently limited to a single modality. the combination of multimodal data but improves the analysis of traffic relations, but the fusing techniques should be sound and dependable in limited circumstances, considering the data-limited tasks and cooperation-forbidden tasks.

4. Predictive Defense and Battlefield Forecasting

Predictive defense aims to obtain foresight over opponents' actions for decision-support or battlefield-forecasting purposes. For decision support, two predictive models for U.S. force deployment and readiness, which consider more than one decade of historical data, have forecasted military events with classification accuracy between 70 percent and 80 percent over several years of holdout data. Forces or units can also be forecast over predefined windows. Switching to real-time considerations, need-to-know dashboarding has contributed to the battlefield-forecasting operations by means of displaying situational-awareness on the side of the commander in the field. In this area, timeliness is of importance.

Silhouette and user-centered design principles guide a risk-indicator dashboard aimed at helping a Combatant Commander, likely in the field, with decisions concerning the cases when the best course of action is Unknown [29-31]. Such decision-support systems require decisional-information quality criteria, a discussion on timelines, a model of decision-matrix quality, and a consideration of α-risk management—foregoing the opportunity to engage the enemy's forces when a losing proposition—along the lines of bias-test considerations of adversarial-predictive models. Research is also ongoing on new risk indicators (e.g. a traffic light) and the combination of U.S. force-posturing decisions with the effects of other influence factors (e.g. intelligence, diplomatic actions).

4.1. Predictive Models for Force Deployment and Readiness

The American army has performed predictive analyses to aid in planning long term and the availability of how to evaluate the readiness of weekends with associated patterns of absenteeism. These analyses are based on (1) simple probabilistic function forecasts of military deployments and (2) a Markov model that generates troop loss rates from the presence of military installations and supports training requirements. NATO deployment predictive models have been adjusted and tested over the recent past but need a development cycle that spans deployments by other combatant commands. Existing and proposed systems support short-decision-cycles and are subject to overloading their users, are pessimistically biased, and under-resource [3,32,33]. Aggregation and interactivity are needed to ensure timely assistance for long-term planning of military-readiness-supporting logistics.

Much of American military intelligence on allied and partner combat-ready forces is carried out on weekends. The nature of the exercise schedules is exposed such that open-source evaluations of places, facilities, preparedness, capabilities, and readiness can be assessed. Recent statistics of deployment of the military, and personnel readiness, have shown the associated trends of absence with potential foresight. Such relationships are being measured through statistical analysis and the model of predicting in effect to ensure that eventually there will be no threat of troop absences to national internal security requirements. One of them is a basic probabilistic models which are applied to predict military deployment under repeated contingency. A second uses Markov to come up with the rate of troop loss based on the availability or unavailability of the military installations as well as to facilitate training requirements.

Although currently focused on NATO countries, the models' foundations can be replicated elsewhere. The deployment forecast for the USAFR weather reconnaissance mission during increased hurricane activity in the eastern United States and on the territory of the Bahamas has proven accurate [4,34-36]. The analysis is awaited to be conducted on the case of the USAFR hurricane timing mission. The 4th of July deployments in aid of American India Week are yet to be developed in terms of predictive probability surfaces that can be applied.

4.2. Real-Time Situational Awareness and Decision Support

Intelligence analysis under conditions of uncertainty and pressure often resembles a series of bets — some large, some small. Adaptive-user knowledge discovery systems designed to assist in situational awareness (situation monitoring and interpretation) and decision support (supporting decisions undertaken in highly dynamic environments) can provide terminal users with easy access to complex information from the state of the army's services and to visual geolocation information presented in a geospatial-based context [37-40]. Such systems present many challenges because information from intelligence sources should be made available in near real time. The information infrastructure must allow users to go from data to information in minutes or at most hours — rather than days.

All such systems must therefore be designed with the end users in mind. Usability tests often reveal a lack of real-world knowledge that impacts how users cope with the information provided. A new set of testing metrics has been proposed that evaluates the success of such a dashboard as a decision-support system—indicators of broader ramifications of intelligence events —and also provides a framework for evaluating the risk of escalation of an event or series of interactions.

5. AI-Driven Logistics and Operational Support

AI is improving logistics that require optimization and is also being directly integrated into intelligence-surveillance-reconnaissance systems that provide information support for military operations. Logistic applications usually involve optimization of resource allocation or transport/supply networks and may consider constraints such as the need to satisfy certain arrival deadlines or span an area in minimum time [4,41,42]. The metrics tend to lay stress on the working throughput of the hidden Markov model of the association of signals and behavior in a maritime setting. In military terms, AI algorithms with competence in tracking, profiling, and predicting behavior of vessels in traffic-prone locations have a great potential of application.. In this case, the implementation of decision support capabilities or monitoring systems with gradual escalation of intervention human supervision increases the reliability of the proposed systems—detecting, profiling, and predicting the behavior of vessels in complex scenarios.

Various concepts of intelligent surveillance systems have been developed in recent years, enabling high autonomy in mission execution—taking decisions on monitoring and possibly qualifying detected events. Emerging technologies for intelligent surveillance will certainly contribute to better performance, but by definition, the human in the loop cannot be neglected. Command and control systems must always be prepared to manage these systems and make the final decision concerning the action to be taken on detected events. Improvement of the overall performance must involve responsive and effective interaction with humans—not only a reduction of the required human effort.

5.1. Logistics Optimization and Resource Allocation

In military success logistics play a very important part. Operation of extremely complicated nature presupposes quick procurement of workforce and equipment which can maintain the huge maneuvers. The allocation and mobility of forces has long been traditionally planned, however, in a non-optimized way. Although the issue of troop readiness has been of immense impact, few machine learning models have been developed to solve it. First efforts were aimed at ensuring reduction in transportation expenses in supply lines.. More recent research used complex event modeling to estimate time-dependent flow of categories of units; the goal was maximizing throughput while satisfying unit flow conditions. An

algorithm driven by reinforcement learning tested the real-time allocation of surface vessels in a joint mobility operation, showing potential for success.

Surveillance logistics are also critical on the battlefield. The forecast of the best position for an unmanned aerial system to cover a mission is very demanding, needing the evaluation of thousands of UAS candidates. A hybrid approach used genetic algorithms to find promising candidates, then Deep Reinforcement Learning reduced the search space. The result is a surveillance UAS's mission profile, given a destination to be monitored in terms of the best coverage quality level. Another planning problem addressed the distribution of unmanned ground systems on a surveillance mission by ground, air, and marine forces. Considering the heterogeneity of the ground units and avoiding conflict, the allocation was converted into a mini-max problem in a flow network with two types of vertices (ground units and communication hubs) plus one dispersed sink. For reinforcement learning, the next state was given by the result of the UAS surveillance assignment.

5.2. Surveillance and Mission Support through AI Systems

Innovations in monitoring capabilities can support intelligence collection and mission execution. These AI systems can assess a variety of monitored sites and events, from space assets to material stockpiles to ground movements to videos of interest. Surveillance methods can range from the sole detection of some phenomenon to a full autonomous-lead capability [43-45]. If decision-makers consider certain developments serious enough, the full escalation protocols need to be applied, including human-in-the-loop control during the most sensitive decisions. The aim is to oversee mission-supporting activities with minimal user intervention and place operators on systems that really require human judgment for successful output.

Societal-scale event detection constitutes a modestly autonomous mission-supporting capability. Events from daily life that have broad human interest are monitored and geolocated using crowd-sourced photo-sharing platforms, Line, or Twitter. Scalability and real-time requirements demand light processing. For visual verification, the full set of photos of concern is processed within a human-in-the-loop mode. Evidence of security-relevant material accumulations may also be monitored using social media, such as YouTube or Weibo. Sufficient video-geotagging can allow the day-to-night transition for temporal context. These detections overall support decision-making about full-scale deployments of costly high-fidelity systems like drones or satellites (or even human observers).

6. Ethics, Oversight, and Risks of AI in Intelligence

Artificial intelligence in US intelligence brings out morality and danger with a focus on openness, discrimination, and counterdangers. Its auditing, responsibility, compliance with the law, and possible discrimination should be examined.

The AI systems should be characterized by transparency and auditability. There should be policy and technical structures which can keep the officials on toes who make automated decisions. Special attention should be paid to signals intelligence (SIGINT) and foreign power threats since the laws restrict the release of the sources. Firms that publish undercover activities should publish their AI platforms. The use of audit trails ought to be proper. Such processes as training data retention, data deletion, and data usage are essential.

The bias of artificial intelligence causes false accusations, discrimination of vendors and attributed intent. Training data may be non-representative or prone to testimony. Adversarial options use the deficiencies in learning to manipulate the data, leading to either the false alarm or the false omission to detect the data. Intentional poisoning that applies defensive strategies based on boundaries only deals with problematic cases, and must be tested and used in untested contexts. Adversarial attacks can be resolved with dynamic training labels, which are picked randomly. Strong sets of rules entail scanty exploitation by indicators of spoofing. Digital watermarking and sensor fusion of adversarial examples as well as distortion detection make AI more reliable.

6.1. Transparency, Accountability, and Legal Considerations

Decision outcomes arising from AI-based analysis affect operational commands. However, civilian oversight does not transparently cover this process (Lentzos, 2020). The White House Office of Management and Budget directs US government agencies to propose privacy and civil liberties safeguards for disclosure. Geospatial intelligence is the most operationally germane area as it requires constant real time data development, and extremely fast decision making. Removal of the human supervision can however be achieved but it comes with dangers of blind trust and interpretation (Michaels et al., 2028). Through democratization of intelligence analysis, geospatial detectives systems which are accessible to the public could be used in detecting crimes. However, misinterpretation or untimely tasking entails the use of decision cycles that are beyond human capacity and risks.

Detection and situational awareness are especially valuable for available system resources. It makes the identification of events difficult due to the temporal and spatial resolution difference caused by large numbers of simultaneous signals [9,46-48]. Although the analysis is backed by background noise suppression based on the unsupervised learning, high false alarm rates are still inalienable to the trained models, which makes the entire method rather questionable (Wang et al., 2023) and require the establishment of new conditions. Ground traffic analysis takes into account prior and subsequent system states, generating complete geo-coordination-based anomaly links in response to unusual main characteristics (Huang et al., 2022) and therefore also allowing for improved operational defence and law enforcement success rates. Graph feature fusion Sensitivity is improved by Graph-based feature fusion using improved separation into non-overlapping noise. Meta-engrained traffic analysis provides uncomplicated situational awareness appraisal with no detection insight, yet because of depending on degree circulation qualities that might possibly not enhance detection aptitude. Limitation of the practical application of tools is constrained by user authority as well as the restrictions on behaviour-connection-data.

The primary weakness of AI is associated with negative adaptation to the changes in data-sets in the domains of model input. Data-poisoning defence and worst case perturbation testing enhances resilience (Zhang and Zhang, 2023).

6.2. Bias, Adversarial Tactics, and Resilience

Adverse side effects (i.e., false flags) caused by feasible attacks on AI systems during their deployment must be avoided, and the systems must also be rendered robust against adversarial tactics during development and testing. False flags due to bias can happen in AI systems when deployed at scale when the phenomenon that entails these in the training data is under- or overrepresented. The events might be used as a zero-day attack by the parties that would be able to creatively modify their SIGINT or network

traffic and provoke disastrous misconduct of AI systems. Detection of potential training data poisoning by adversaries that are aware of the AI system under consideration could also be used to defeat AI systems during their deployment. Techniques to limit and detect such vulnerabilities are therefore critical.

AI systems used in intelligence should not only provide accurate probabilistic predictions on their anticipated region of operation, but response times also need to be fast enough for the relevant situation. Moreover, the required indicator of behavior change should be defined, and the information displayed emphasized during the design of the supporting technology, to facilitate operator trust and confidence. The dashboard should also support the monitoring of uncertain KPI values and correlate them with the observed situation to assess the reliability of the risk indicated. Such indicators must be regularly re-evaluated, as conditions dependency may result in completely different meanings.

7. Conclusions

The extensive use of AI in the context of GEOINT and SIGINT-based intelligence gathering and analysis helps to increase efficiency and decrease response time, as well as better operational decision-making. The systems of interpreting satellite imagery and the trends of life of targets, based on AI, help to collect faster and more efficiently. However, the created capabilities also bring new burdens to the US decision cycle and initiate the combat of risks that are related to excessive reliance or misunderstanding of the AI-generated products. The danger of the adversarial strategy of hardware and software cannot be underestimated; on the contrary, it has to be included in the development of the adversarial detection systems and the general use of AI in the context of GEOINT.

AI capabilities in visual analysis and pattern-of-life modeling rely on analysis and decision support systems developed for the detection and depiction of object-class abnormal activity. Other such tools use the data of operations to determine whether any force is ready to attack and whether it is sustainable enough to sustain the operations until it is detected or surrendered. Some combination of forces has been proven by history to be taken correctly with these predictive models. Horizon-scanning and foresight tools are still in development in the areas of missions. These models will guide the leaders in picturing the impact of altered realities on operational planning and execution, and the degree of the risk in case the deficiency in time, forces, or sustainment arise.

References

[1] Bevilacqua S, Masárová J, Perotti FA, Ferraris A. Enhancing top managers' leadership with artificial intelligence: insights from a systematic literature review. Review of Managerial Science. 2025 Jan 22:1-37.

[2] Mohapatra PS. Artificial intelligence and machine learning for test engineers: concepts in software quality assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:17.

[3] Bolton C, Machová V, Kovacova M, Valaskova K. The power of human–machine collaboration: Artificial intelligence, business automation, and the smart economy. Economics, Management, and Financial Markets. 2018 Dec 1;13(4):51-6.

[4] Brynjolfsson E, Mcafee AN. The business of artificial intelligence. Harvard business review. 2017 Jul 18;7(1):1-2.

[5] Canhoto AI, Clear F. Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. Business Horizons. 2020 Mar 1;63(2):183-93.

[6] Chen L, Jiang M, Jia F, Liu G. Artificial intelligence adoption in business-to-business marketing: toward a conceptual framework. Journal of Business & Industrial Marketing. 2022 Apr 15;37(5):1025-44.
[7] Chen Y, Biswas MI, Talukder MS. The role of artificial intelligence in effective business operations during COVID-19. International Journal of Emerging Markets. 2023 Dec 12;18(12):6368-87.
[8] Chu SC, Yim MY, Mundel J. Artificial intelligence, virtual and augmented reality, social media, online reviews, and influencers: a review of how service businesses use promotional devices and future research directions. International Journal of Advertising. 2025 Jul 4;44(5):798-828.
[9] Dirican C. The impacts of robotics, artificial intelligence on business and economics. Procedia-social and behavioral sciences. 2015 Jul 3;195:564-73.
[10] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[11] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.
[12] Enholm IM, Papagiannidis E, Mikalef P, Krogstie J. Artificial intelligence and business value: A literature review. Information systems frontiers. 2022 Oct;24(5):1709-34.
[13] Feuerriegel S, Shrestha YR, von Krogh G, Zhang C. Bringing artificial intelligence to business management. Nature Machine Intelligence. 2022 Jul;4(7):611-3.
[14] Han R, Lam HK, Zhan Y, Wang Y, Dwivedi YK, Tan KH. Artificial intelligence in business-to-business marketing: a bibliometric analysis of current research status, development and future directions. Industrial Management & Data Systems. 2021 Nov 10;121(12):2467-97.
[15] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.
[16] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[17] Kim H, So KK, Shin S, Li J. Artificial intelligence in hospitality and tourism: Insights from industry practices, research literature, and expert opinions. Journal of Hospitality & Tourism Research. 2025 Feb;49(2):366-85.
[18] Kumar D, Ratten V. Artificial intelligence and family businesses: a systematic literature review. Journal of Family Business Management. 2025 Apr 17;15(2):373-92.
[19] López-Solís O, Luzuriaga-Jaramillo A, Bedoya-Jara M, Naranjo-Santamaría J, Bonilla-Jurado D, Acosta-Vargas P. Effect of generative artificial intelligence on strategic decision-making in entrepreneurial business initiatives: A systematic literature review. Administrative Sciences. 2025 Feb 18;15(2):66.
[20] Loureiro SM, Guerreiro J, Tussyadiah I. Artificial intelligence in business: State of the art and future research agenda. Journal of business research. 2021 May 1;129:911-26.
[21] Maiti M, Kayal P, Vujko A. A study on ethical implications of artificial intelligence adoption in business: challenges and best practices. Future Business Journal. 2025 Mar 13;11(1):34.
[22] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[23] Mumtaz S, Carmichael J, Weiss M, Nimon-Peters A. Ethical use of artificial intelligence based tools in higher education: are future business leaders ready?. Education and Information Technologies. 2025 Apr;30(6):7293-319.
[24] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.
[25] Naim A. Role of artificial intelligence in business risk management. American Journal of Business Management, Economics, and Banking. 2022 Jun;1:55-66.

[26] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.
[27] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..
[28] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.
[29] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.
[30] Panda S. Scalable Artificial Intelligence Systems: Cloud-Native, Edge-AI, MLOps, and Governance for Real-World Deployment. Deep Science Publishing; 2025 Jul 28.
[31] Perifanis NA, Kitsios F. Investigating the influence of artificial intelligence on business value in the digital era of strategy: A literature review. Information. 2023 Feb 2;14(2):85.
[32] Porkodi S, Cedro TL. The ethical role of generative artificial intelligence in modern HR decision-making: A systematic literature review. European Journal of Business and Management Research. 2025 Jan 23;10(1):44-55.
[33] Qin C, Zhang L, Cheng Y, Zha R, Shen D, Zhang Q, Chen X, Sun Y, Zhu C, Zhu H, Xiong H. A comprehensive survey of artificial intelligence techniques for talent analytics. Proceedings of the IEEE. 2025 Jun 6.
[34] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.
[35] Quan XI, Sanderson J. Understanding the artificial intelligence business ecosystem. IEEE Engineering Management Review. 2018 Nov 20;46(4):22-5.
[36] Ruiz-Real JL, Uribe-Toril J, Arriaza Torres JA, de Pablo Valenciano J. Artificial intelligence in business and economics research: Trends and future. Business Economics and Management (JBEM). 2021;22(1):98-117.
[37] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[38] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.
[39] Secundo G, Spilotro C, Gast J, Corvello V. The transformative power of artificial intelligence within innovation ecosystems: a review and a conceptual framework. Review of Managerial Science. 2025 Sep;19(9):2697-728.
[40] Sestino A, De Mauro A. Leveraging artificial intelligence in business: Implications, applications and methods. Technology Analysis & Strategic Management. 2022 Jan 2;34(1):16-29.
[41] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[42] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.
[43] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.
[44] Singh N, Chouhan SS. Role of artificial intelligence for development of intelligent business systems. In2021 IEEE International Symposium on Smart Electronic Systems (iSES) 2021 Dec 18 (pp. 373-377). IEEE.
[45] Soni N, Sharma EK, Singh N, Kapoor A. Artificial intelligence in business: from research and innovation to market deployment. Procedia Computer Science. 2020 Jan 1;167:2200-10.
[46] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.
[47] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.

[48] Shivadekar S, Mangalagiri J, Nguyen P, Chapman D, Halem M, Gite R. An intelligent parallel distributed streaming framework for near real-time science sensors and high-resolution medical images. In50th International conference on parallel processing workshop 2021 Aug 9 (pp. 1-9).

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 7: Ethical Challenges of Autonomous Weapons and AI in Modern Warfare

1. Introduction

Discussion of the modern warfare can now heavily rely on the ethical issues and political issues of the emerging technologies, including Lethal Autonomous Weapon Systems (LAWS), drone swarms, and AI war decision-making. Although everyone agrees on the basis that LAWS are weapons that can, once deployed, act independent of the sensor information to choose and attack the targets, various definitions have been given depending on the scope of the autonomous functions. These two directions have become two key dimensions of classification; (1) according to the type of tasks allocated or the tasks to be delegated include situational awareness, target selection, and engagement, and other stages of the decide-attack-assess-detect-exploit cycle) and (2) according to the degree of delegation between human-in-the-loop and human-out-of-the-loop to fully autonomous systems. The identification of scope and degree of autonomy is used in both technological research and testing as well as ethically and legally in LAWS (proportionality, discrimination, civilian protection) and the extent of dehumanization of the use of force.

Although the humanitarian and ethics-oriented argument in favour of the banning of LAWS is very strong, the counterargument has a tendency of highlighting operational gains made or anticipated, without given proper treatment of the security losses or ethical tradeoffs.. Security cannot be treated in isolation, for harm to innocent civilians is the most potent criticism that can be levelled against military action. To ensure ethical functioning within diminished or non-existent ethical guidance during combat action, LAWS will thus require continuous, responsible human oversight and possible intervention during actual military operations. Examination of the ethical and legal issues invokes verification and compliance questions, and, more generally, the readiness of established frameworks for these new developments.

2. Lethal Autonomous Weapon Systems (LAWS)

Lethal Autonomous Weapon Systems (LAWS) are weapons systems that on activation can automatically target and hit their targets without the driver having any influence over the system. With a higher autonomy level, weapon control becomes looser and even the human role changes to the pulling of the trigger and taking a prescribed behavior or task. LAWS can have advanced sensor fusion ability enabling them to fire upon a series of targets without performance reduction. These systems are therefore characterized by a reduced decision cycle of the human operated platforms and as such, can be engaged faster [1,2] . This way, LAWS can be seen to blur the kill chain. Nevertheless, the kill chain does not lose human agency but just loses its connection with decision making.

There remains an open debate on whether LAWS violate key principles of humanitarian law (i.e., proportionality, discrimination, humanity). An ethical concern continues to be that, by removing human agency from the kill chain, LAWS are forbidden from executing orders that may contravene international norms or law. Their operation is also of a dehumanizing nature, detaching the "human suffering" dimension from conducting an operation involving loss of human lives. ethal Autonomous Weapon Systems (LAWS) are weapons systems that on activation can automatically target and hit their targets without the driver having any influence over the system. With a higher autonomy level, weapon control becomes looser and even the human role changes to the pulling of the trigger and taking a prescribed behavior or task..

2.1. Definitions and scope

LAWS can have advanced sensor fusion ability enabling them to fire upon a series of targets without performance reduction. These systems are therefore characterized by a reduced decision cycle of the human operated platforms and as such, can be engaged faster. This way, LAWS can be seen to blur the kill chain. Nevertheless, the kill chain does not lose human agency but just loses its connection with decision making. LAWS can be classified based on the degree of anthropomorphism, the distribution of sensory data-processing, the duration of the decision cycle, the concentration or dispersal of the kill-chain capacities, and their intended purpose.

The most relevant ethical concern regarding the use of these high-capacity machines would be whether LAWS would comply with the principles of distinction, proportionality, and military necessity, which are a set of poles of the International Humanitarian Law [3-5]. Any deviation from compliance with these canons would trigger a revival of the deontic argument that holds that the state of acting wrongfully lies in the type of act itself rather than in its foreseen consequences. The potential ethical costs of LAWS go beyond their likely tendency to lay to waste the distinction between combatants and civilians. The highly adaptive, flexible, and dynamic character of war and the fact that humans have employed pre-programmed machines as large-scale force multipliers that introduce ever-new speeds and types of tempo in operations in the presence of man-in-the-loop decision-making also raises questions about the dehumanization of war through the fading presence of human descriptive, predictive, and prescriptive-accommodation faculties offensively and defensively, and, even more problematically, through their complete removal. LAWS can therefore be understood in part as a possible rupture and in part as a mere extension of pre-existing events in the field of armed conflict.

2.2. Ethical implications

Implementation of the Lethal Autonomous Weapons Systems brings a number of ethical issues. One of the strengths compared to other systems is that such systems claim to accommodate the requirements of proportionality and discrimination without the involvement of humans. Proportionality is the requirement to make sure that the collateral damages are not disproportionate to the expected military advantage. The duty to draw the line between military objectives and the protected persons or property is the discrimination. In case these requirements are met, the armed forces fighting in the combat area can be supported with the help of automatic systems. Nevertheless, the lack of human authorization to the kill decision compromises one of the most important aspects of the human justification of the drone killings, which is that of human lives [6,7]. The absence of a human operator of the kill decision raises once again the issue of the deontological objections of the very killing of the enemy. The fact that there can be such thing the dehumanized war has shown the fear that the old time restraints used in conflicts like the

issuance of shame or moral outrage are being washed away through the physical and emotional thereof offered by LAWS. The deontological objection lies in a simple question: Can killer robots, once they fulfill the obligations of proportionality and discrimination, be deployed to kill other humans? The deterrent capacity of a state's possession of LAWS may indicate its inability to master dehumanized war.

The humanitarian risks of LAWS stem not from their expected military effectiveness but from the possibility of a malfunction resulting in civilian harm [2,8-10]. Several forms of malfunction have been identified, including 'man-in-the-middle' attacks, adversarial exploits, swarming, and hostile environment. Commander responsibility for civilian harm caused by LAWS must be seen against this background. IHL requires that all feasible precautions must be taken to minimize civilian harm. Two examples emerge. First, the need for redundant safeguards in the design and operation of LAWS that are able to breach hostile environments. Second, that the assignment of duties to LAWS must be within the limits of the knowledge of the commander and the risk involved.

2.3. Strategic and humanitarian considerations

The deterrence and escalation activity is a strategic evaluation of LAWS, whereas the potential risk of malfunction, normative compliance, and the potential dehumanization of warfare refer to the humanitarian consideration. The possibility of the LAWS to promote the deterrence process within an adversarial relationship is mitigated by the possibility of unintended escalation during a crisis situation. When non-human actors start assault, there is a more significant threat of a local conflict with international consequences. Deterrence would nevertheless be strengthened if an improved ability to target military objectives in-depth quickly and at an acceptable level of collateral damage also discouraged military attacks by an adversary.

Specific missions in the form of LAWS provide the potential of an operational gain but eliminate the human factor, which cannot be jeopardized [1,11-12]. Other military forces outside the United States and China are enhancing the LAWS on limited operations like the protection of the maritime ports or air base. The economic rationale behind the multiplication of military forces with the LAWS will be based on the efficient presentation of the increased reliability rates, strength and stability of the LAWS armed forces. The possibility of the legality of LAWS and concern of the humanitarian implications have not been reduced yet. Even with the most rosely colored assumptions, this would make humanitarian law unenforceable because of the fact that it would be impossible to check whether compliance with the principle of distinction and proportionality has been checked.. The technological sophistication of LAWS under development will reduce the likelihood of malfunctions, but a greater complexity–capability trade-off may increase their susceptibility to adversarial exploitation and abuse. Furthermore, the gradual transfer of war-making capabilities and responsibilities away from human agents towards LAWS raises the risk of dehumanizing warfare.

3. Drone Swarms and AI Combat Decision-Making

Swarms of drones Drones with owner drones have become a promising form of military technology. Massive swarming enhances the utility of UAVs, taking advantage of their main features: low prices, size, expendability, sensor distance, and the ability to collect information. Favourable developments in miniaturization, MEMS systems, swarming algorithms, cloud computing and electronic power / flight time have led to new capabilities in the military [13-15]. Although present day demonstrations use a maximum of ten drones, there have been proposals of large tactical uses of swarms in swarm to swarm

engagements and this indicates that future swarms will have hundreds or thousands of units. As this growth occurs, the coordination and communication of the swarm operation becomes more complicated; quick coordination might rely more on swarm collective cognition than centralized coordination or any agent-agent negotiation. Like federated deep reinforcement learning, technologies are in place to enable effective and robust multiclass drone swarming. Nonetheless, the need of commands of resilient heavily redundant input sources to spatial awareness and navigation is also demonstrated by large-scale swarming.

There are concerns regarding the deployment of AI decision-making into battle and other potentially lethal versions, especially in a wider team of humans like conducting operational plans and performing actions [16,17]. The lack of accountability between the commanders who give the approval and command AI agents, between the AI agents, who choose and act without human intervention, and between the non-state actors, who have no unvalued target, no command-and-control center, and no command-and-control chain is especially crucial, and so are the risks of civilian casualties and capacity to adhere to international law. In addition to simple proportionality, just war is complex due to a comparative ability of small numbers of AI to make the hostility or terrorism cheap or make the sanction cost of using indiscriminate force much less against a superior enemy compared to innocent bystanders.

3.1. Technological capabilities and limits

Swarms of drone robots that the AI incorporation in robotic systems made possible are a new element that opens up new ways of fighting the war. Recent massive projects like the LOCUST (Low-Cost UAV Swarms support Persistent Maritime Superiority) program of the U.S. military and the Gremlins program of DARPA have shown that autonomous swarming and drone swarm support can be viable. These innovations are not developing fast; the principle of functioning of such systems is not completely elaborated, and the majority of logistics and support activities are not being robotized. Introduction of AI in the drone swarm will undoubtedly increase their efficiency in such areas to both the swarming drones and the pantheon of support operations. A number of technological abilities and restrictions are still in play regarding ethical and legal debates. Swarms operate very much parallel operations in serving a tactical purpose. Drones that are possible with the assistance of AI decision support may make a discreet mission to survey a wide territory and assist in dense attacks, which will complicate force protection of the target. Besides, these operations may need little to no input of a human commander, being fully reliant on drone-to-drone communication as a very robust communication framework. It is also, however, true that the control systems that allow these functions are weak; the enemy forces can destroy the drone networks by means of information warfare [12,18-20]. Practically it is common knowledge that as soon as there is a dependable C2 structure, non-swarming solutions will tend to be more efficient and robust. urown swarins. AI and drone swarms are now addressing the next level of problem; cooperation and coordination forms that radically alter communications and force presence.

Such communication-based factors can radically influence the considerations normally applicable to drone-based operations, as key elements of doctrine and control structures can become fundamentally altered compared to the current baseline application of technologies. As a ramification, information warfare can gain an even greater role and density in future operational settings. Despite further attrition of normal C2 and information support roles, redundancy of networks and paths for information transit can reduce decision latency at the level of swarming actions. Compared to a standard non-swarming drone capability, the net impact on risk might remain relatively constant; although standard drone capabilities

can attain, verify, and act on decisions faster, swarm capabilities can exploit coherent decisions made by other actors without requiring that a command pathway remain properly functioning and reversible [21-23].

3.2. Ethical and legal concerns

Drone swarms also provoke ethical issues of agency, accountability, proportionality, and discrimination that are vividly brought to the fore due to the possibility of the lack of the human commander-in-the-loop. The use of swarming drones also stretches the limits of the International Humanitarian Law especially as far as civilian protection is concerned. As fast and reduced-onboard information processing capabilities may represent considerable benefits in the context of air defense against the missiles or conventional-based forces, the combination of pace and recidivism demonstrates an unparalleled problem of adherence to current norms.

The arguments for and against a human-in-or-out-of-the-loop delegation of authority are well-established, as are the asymmetrical issues that arise in hybrid conflicts. Three considerations stand out. First, whether LI/LAWS operators act as the creators, constructors, or merely manipulators of the deciding action affects the moral agency behind the decision. Second, whether the delay introduced by a commander-in-the-loop (or alternatively, an approving authority) is disproportionate to the possibility of preventing a mistake is an empathy-based reflection that goes beyond pure speed advantages, for it weighs the speed advantages for one side against human supervision for the other. Lastly, the autonomy level also correlates with chances of malfunction along with the swarming, introducing the extra layer of complexity: any attack on a target without high importance is bound to be successful but wrong, and system pressures are likely to show an attack out of a robust redundancy and toward aggression, further increasing the number of unintentional reactions to the perceived threats [24,25].

3.3. Operational implications for doctrine and risk

The military operation doctrine is conditioned by positioning aspects which influence military operational functioning. Human man-made platform and its entrusted operator operational safety doctrines put lesser faith in technology to evade interpersonal relationships, socio-complexity of the warfighting environment, and human foibles. The building of such doctrines with human-on-human operation safety does not give that of autonomous entities. The driverless vehicles require mechanical separation in order to be safely operated. Increased speed is advantageous with autonomy. Current efforts in combat enhancements put the human-on-machine latency in decision making in a negative picture that is now enshrined in the phases of steps that are now a part of operations that are detect, decide, deliver and see. The more operational execution speed distances devices to the physical limits of interacting with humans, the less understandable are the decisions being provided by machines. Multi-domain operations introduce additional player forces which have less supervision but have to be fault tolerant as adversary tactics develop to interrupt own-side swarming operations. The structures of command and control will also be required to be checked and revised to provide own military safety.

Two underlying operational principles appear The first one is redundancy within the force structures, which is a guarantee of operational safety; the systems cannot be left to operate completely without supervision [26-28].. The ability for enemy force protection against these autopilot-enhanced drones will determine when the threat will be credible saved by the woke naysayers' crying that they are a threat to

civilians. Echoing the effect of (nuclear) credible threat behind mutually assured destruction, the crying of the naysayers will lessen the necessity for perceived enemy protection.

4. Human-in-the-Loop versus Human-out-of-the-Loop

One way this difference traces the level of human decision making powers and the ability to control lethal attacks. Although the definitions of the US department of defense concern mainly a theoretical understanding of the so called delegation to the machines, the LAWS as a category of weapon systems can be categorized within a spectrum of control by man with distinct practical implication.. While not all decision-making can be delegated to computers, aspects of computing speed and analysis can be incorporated into the decision-making cycle for particular attack decisions. The distinctions highlight that allocating responsibility and understanding accountability issues in future conflicts will be much more complex if weapons relying on human-out-of-the-loop decision processes are present.

The distinction examines the location of moral agency across a military operation that can encompass combat actions but also the supporting actions involved in preparing for conflict. With human-out-of-the-loop systems, the user determines targets but does not verify situations or target profiles before attack decisions are made [29-31]. Additionally, these systems are designed to operate independently of the user's intentions during the attack phase. Thus, any attack decision is conducted without human verification or involvement and based on automated target profiles for a non-human recognisable target without the user's verification. The distinction characterises existing weapon systems using human-on-the-loop for target selection, monitoring, control, and termination, and identifies future systems operating on a human-out-of-the-loop basis.

4.1. Conceptual distinctions and practical realities

The term human-in-the-loop is used to describe the situations, when the final decision to engage or not to engage a target to a target purpose is taken by humans, and human-out-of-the-loop approach describes those situations, when automated systems receive the authority to engage, but human factors are replaced as parts of the system to engage a target [3,32,33]. It is important, however, to not treat these two concepts as framed in binary terms. Between both of these extremes exist many forms of delegation of decision authority, ranging along a continuum from humans making the decision with specific assistive recommendation inputs on one end to the automaton making deciding with pursuit of directive objectives provided by humans on the opposing endpoint. In-between these poles, varying levels and modes of human oversight and lateral interaction exist for an unbroken chain of human decision-making authority, either through incremental suggestion and permission gates or limiting latencies.

An additional critical elucidation is that the human-in-the-loop framing is more appropriate to decision cycles of particular components of the kill chain. The human-in-the-loop feature of the LAWS operational cycle can sometimes in fact be rather a constraining behavioral layer between options of operational risk-tolerance, and not necessarily of operational speed. LAWS may bring about a visible speed bonus, but to the effect of diminishing the comparatively time-sensitive and perceivable scenario to tolerable risk levels. Until the challenge-full aspect is sufficiently curtailed, human militaries can continue to act faster than LAWS, except in certain cases involving very high levels of concurrency to which the LAWS can deploy combinatorial flooding. Whether LAWS are able to fulfill the decisive role of extreme speed as assigned in doctrine is thus in practice left conditioned by the state of the adversarial situation until the interacting conditions are verified to be compliant with the risk-tolerance algorithm of the LAWS.

Added complexity on the conditions of decision in transit is the competing execution standards, which affect aspects in all stages of the kill chain [4,34-36]. The intrinsic conflict between operational risk and decisional speed is usually exacerbated by the above-mentioned latency in decision transits attribution to LAWS. The inability for humans to rapidly command-act or even redirect in the face of malfunctions, misconceptions, incorrectly sensed situations which result in targeting errors, or confusion of friend with enemy becomes fundamental constraints.

4.2. Moral agency, responsibility, and oversight

Being responsible for making a decision means being accountable for that decision, and such accountability is the bedrock of morality. This is the notion of moral agency that is used in the debates regarding the moral attributes of AI and autonomous weapons. The questions of attribution of wrongdoing preventative and minimization of collateral damage and the fairness of moral ethical conduct are what revolve around the actions of the person who can be held responsible with regard to the making of decisions that underplay the perception of legitimate authority [37-40]. The answers to such questions are also a matter of operational performance. Treating autonomous weapons as only tools—like an airplane or a tank—seems to minimize the associated problems, but is rarely entirely valid. Such a view ignores not only how tools can be misused but also the importance of an ethical component to the actions in the decision chain. A reasonable question, accordingly, is whether the doctrine and command structures of using autonomous weapons and autonomous decision-making should more or less change these aspects.

Some lack of functionality in operating LAWS may ease the accountability problem, but in the case of human-in-the-loop (HITL), human-on-the-loop (HOTL), and human-out-of-the-loop (HOOTL) capabilities, the abstraction makes the assignment of guilt and punishment more difficult. Particularly in the HOTL and HOOTL stages, the decision-making process moves away from being an action of a specific person or group that can be assigned wrongdoing. Such lack of a direct person also raises questions about supervision, with the assurance mechanism also being somewhat automatically in the hands of the developer. In particular, HITL in cases in which the operator does not speak the language of the actors or environment is, in any decision cycle, mentally working under a different language in that specific and critical stage; even using a language translator. The probability of antiambition simulator attack is also higher when lethal forces are ship or air-born and are of HWOT type.

4.3. Operational reliability and decision latency

Speed advantages are a crucial operational benefit of AI-based combat decision-making. Machines are capable of assessing and reacting to situations far more rapidly than humans. Such capabilities are significantly accelerated when the combat decisions of large swarms of weapons are linked and advanced together [4,41,42]. Nevertheless, for these benefits to be brought to fruition, movement in the early stages of a conflict—and indeed for its duration—will not only require that humans remain comprehensively ‘in the loop’, but also that their control over decision-making remains sharp and strong. Without this early movement into the lower regions of the decision tree, the military advantages of speed and thorough, really mission-compliant, decision-making are lost. At the other end of the cybernetic loop, however, questions remain over the nature and effect of the inevitable qualitative and temporal advantages that AI-controlled weapons have over non-swarms or human-operated forces. For swarms, decisions need to be verified before the firing messages can be sent. However, technology currently does not permit human

verification of very large and/or massively parallel, group decisions because it will take human minds too long to fully comprehend the decision.

Latency will always impose a burden on human-control-mode questions. Pressure will build for delegating control to the swarm. While the speed advantage of machines will permit greater tolerance by human operators for mistakes, swarms are not going to be allowed to make strategic decisions that affect the outcome of a war. Swarms will act to protect themselves equally quickly. Should one of them become malfunctioning, the defect will be systematically examined, redundancies will neutralize its pernicious action, and a cause will be found [43-45]. Humanoid AI can be readily designed to think tactically and yet without empathy or understanding of deception. The constraint of holding-up-decision commands will thus be critical.

5. Accountability, Treaties, and International Norms

From a legal perspective, the introduction of the most formidable autonomous weapons raises questions of accountability and criminal responsibility when a state of war is employed. Of particular interest is ascription of criminal responsibility to the states that manufacture or use these weapons when these systems are not thwarted. In international relations, states are regarded as the principal actors. As a result, discussions about war crimes or crimes against peace concentrate on explaining these events and the actors responsible for them. However, only a state can be labelled a “war criminal” and thus be punished. The question arises often as to whether non-state actors such as terrorists or armed groups can be said to have committed a crime against peace by initiating hostilities against a state. The concept of “crime against peace” was coined in an era in which the existence of the state was paramount, and it was understood that internal conflicts were simply an expression of an attempt to protect the status of both the state and the social group that dominated the political system in that state.

When the system used by non-state actors to wage war against a state includes, for example, the use of drone swarms equipped with LAWS and acts of sabotage against the deployed C4ISTAR systems of the state, these state interests as expressed in the right to self-defense shape the political discourse justifying violations of international law. The Assassination of UBL series of military actions can be interpreted as actions combining unilateral political action with military action taken for maintaining the interest of the United States in the existence of the state of Pakistan, although the government of that state had officially condemned these actions.

5.1. Attribution and responsibility in autonomous engagements

Inself-directed interactions, attribution and responsibility is not easily determined by the level of autonomy and changing position of human operators. The state and non-state actors must have clear lines of responsibility to appropriately apportion blame and hold them accountable. Failure of LAWS will cost civilian lives among other illegal implications and makes one wonder whether states have a role to play concerning the consequences. Also, a process of dilution of command chain and absence of human supervision in an operation undermine responsibility. It is important to ensure that such problems are dealt with to reduce misconduct and its effects.

The Lethal Autonomous Weapon Systems (LAWS) and the AI functions are also vital facilitators of the war in the future that can have effects on strategy and doctrine. The way the military should act will change because of the operational adoption of the LAWS and drone swarms of ground, air and naval corps. Before long, the decision of tactical and operational scouting and deployment will be distributed on

the battlefield, imposed in redundancy swarms will allow results in real time communication and prompt responses to the enemy forces before computer combat decisions go to create decisive advantage at the operational and strategic level. The realization of this capability elicits significant changes in doctrine especially on applicability of IHL and command structures required to it.. Yet, the combination may also present escalation risks, undesired consequences in classified missions, rapid-mash-up defence doctrine ruin, and force protection without additional external deterrence [9,46-48].

5.2. Current treaties, gaps, and enforcement challenges

A legal challenge on autonomous weapon systems is the extent of control that human actors will need in the international treaties to be applicable to the case of the Geneva Conventions, the Hague Conventions, and the Convention on Certain Conventional Weapons. Such treaties are indeed effective in attempting to control some form of minimum of violence that war entails.. If autonomous weapon systems are used and engage in armed conflict without that minimum threshold of violence, then the conventions would not apply. In such situations, state or non-state actors using these systems could harm civilians and/or civilian infrastructures, enhancing the exploitation of asymmetric power provided by such systems. Furthermore, when used by terrorist organizations, deterrence and retaliation would be nearly impossible.

On the other hand, the lack of thresholds in the treaties could pose a further danger. factually, the transfer of control to autonomous weapon systems in most cases tends to insist on the need to have a clear but pragmatic definition so that the action is attributable, something that in its turn prompts operational issues. Attribution is an important aspect of making the identified individual responsible in a crime as crimes most of the times utilize the chain-of-command dynamics particularly in matters of future wars where the force is made up of a coalition of states or a series of state and non-state forces.

5.3. Proposals for governance, norms, and verification

Multi-layered regimes promoting innovation need to be complemented by confidence-building measures, necessary for collective and security and ethical accountability in major power relations. Field policies Command-and-control, i.e., connectivity requirements, kill-chain requirements, fire-superiority requirements, interoperable personnel structures, etc., will go a long way to satisfy such demands. Nevertheless, in the absence of formal treaties, external verification is essential in ensuring that deterrence is effective because it is not possible to conduct transparency tests on threshold technologies.

Steps along the road to establishing governance norms are also important at this stage. As it currently stands, the issues of governing LAWS are mostly reduced to the potentiality of a prohibition or a regulation of operation as Germany continues to build the pressure to get a conclusive meeting that would talk about the possibility of a treaty in the year 2023. The more complicated structural issue is balancing between innovativeness and adequate levels of ethical surveillance, verification and enforcement, particularly, or arguments of collective and individual security [49-51]. This can make the existing treaties ineffective or insufficient; the dynamics in alliance can be an innovation in directions that give less priority to humanitarian concern. Any multilayered regime must be well armed. Two applications of the word accountability apply in this scenario: to guarantee that failures of the IHL during resumption or escalation of hostilities are duly evaluated - and that those who perpetrate them should be prosecuted, wherever possible - and to ensure that in less-operationally coupled settings actors are always under tight control of weapon systems and underlying algorithms.

6. Synthesis: Balancing Innovation, Security, and Ethics

The above analyses reveal that there are many ethical issues related to the creation and future deployment of advanced autonomous weapon systems, such as fully lethal autonomous weapon systems (LAWS) that can be used as fully autonomous in the kill chain; swarms of armed drones cooperating with one another in the engagement process; and drones and missiles systems capable of making combat-related decisions with minimal or no human involvement in the decision-making process. The first series of deliberations is whether these liberties can be made to coexist with the set of legal values, which is a delicate balance of operational military needs and humanitarian values that should shape the theory and practice of war. International negotiations in the future shall have to consider the loopholes in the agreements that are committed today as well as the sufficiency of the ordinances and the principles and the verification structures.

Innovation and deterrence are connected, so one of the central strategic evaluations is the connection of the two. Although technological innovations have traditionally been considered as one of the forces of military deterrence, there are serious ethical and collective issues regarding the implementation of decision-making architectures with little human intervention in them. The inherent drive to the military organizations to benefit themselves by systematically increasing their mission and operational benefits, leads to more and more simplified systems, working continually at the border of military reliability. Such tensions mould into a set of normative considerations whose reminder is that effectiveness is not a sufficient factor that should determine whether a technology should be utilised. In addition to a cost/benefit analysis, the human and military community have to face the ethical consequences of uncoupling some responsibility out of military decision making.. Future armed conflicts will be informed and shaped by these control dilemmas—not in less than lethal control, but in how the community of nations constructs a military ethical framework.

7. Conclusion

Striking a balance among a plethora of competing factors affecting LAWS, drone swarms, and AI-based combat decision-making is extremely difficult. On the one hand, the considerable promise of innovation motivating development remains. On the other hand, the potential security concerns and ethical issues resisting such advances are equally strong. To help identify a suitable pathway for these capabilities, an analysis of key considerations surrounding LAWS, drone swarms, and AI-driven decision-making has been presented. Security and ethical principles should serve as a two-part filter guiding the advancement of these technologies, with burgeoning innovation facilitating their development only when no important ethical or security violations are likely. Without this two-pronged process, LAWS, drone swarms, and AI decision-making likely contain very dangerous security implications that cannot be easily alleviated, yet the military, political, and civilian nature of their decision-making processes is such that lack of sufficient acknowledgement and governance may easily lead to cataclysmic consequences. Absent such a process, the integration of autonomy into military actions will most probably lead to a greater future risk than security benefit.

The extensive introductory examination of LAWS, their implications for war-fighting, and military and international law is more than warranted, for if their use is not prohibited, an ethical justification at least must exist. An examination of LAWS, drone swarms, and AI combat decision-making applies standard assistant-commander, officer-commander, and chain of command principles neatly exploited in a military context: when considering these three forms of AI, questions of what they really are can be framed around

the nature and degree of the human role in decision-making. That is, are humans in the loop, on the loop but not in it, or completely omitted? Understanding the practical operation and ethics of such distinctions consequently enhances understanding of the unique capability of human commanders to synthesize qualitative data in creative, responsive, appropriately validated/private/completely secure, and jurisdictionally bound ways, while simultaneously allowing the suitability of machine assistance in achieving increased mission success

References

[1] Zulaikha S, Mohamed H, Kurniawati M, Rusgianto S, Rusmita SA. Customer predictive analytics using artificial intelligence. The Singapore Economic Review. 2025 Jun 6;70(04):1009-20.

[2] Wright SA, Schultz AE. The rising tide of artificial intelligence and business automation: Developing an ethical framework. Business Horizons. 2018 Nov 1;61(6):823-32.

[3] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.

[4] William P, Panicker A, Falah A, Hussain A, Shrivastava A, Khan AK. The Emergence of Artificial Intelligence and Machine Learning in Contemporary Business Management. In2023 4th International Conference on Computation, Automation and Knowledge Management (ICCAKM) 2023 Dec 12 (pp. 1-6). IEEE.

[5] Wang X, Lin X, Shao B. How does artificial intelligence create business agility? Evidence from chatbots. International journal of information management. 2022 Oct 1;66:102535.

[6] Swan M. Blockchain for business: Next-generation enterprise artificial intelligence systems. InAdvances in computers 2018 Jan 1 (Vol. 111, pp. 121-162). Elsevier.

[7] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.

[8] Soni N, Sharma EK, Singh N, Kapoor A. Impact of artificial intelligence on businesses: from research, innovation, market deployment to future shifts in business models. arXiv preprint arXiv:1905.02092. 2019 May 3.

[9] Sollosy M, McInerney M. Artificial intelligence and business education: What should be taught. The International Journal of Management Education. 2022 Nov 1;20(3):100720.

[10] Sipola J, Saunila M, Ukko J. Adopting artificial intelligence in sustainable business. Journal of Cleaner Production. 2023 Nov 10;426:139197.

[11] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.

[12] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.

[13] Saxena M, Mishra DK. Artificial intelligence: the way ahead for employee engagement in corporate India. Global Knowledge, Memory and Communication. 2025 Jan 13;74(1/2):111-27.

[14] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.

[15] Mohapatra PS. Artificial intelligence and machine learning for test engineers: concepts in software quality assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:17.

[16] Reim W, Åström J, Eriksson O. Implementation of artificial intelligence (AI): a roadmap for business model innovation. Ai. 2020 May 3;1(2):11.

[17] Rane NL, Paramesha M, Choudhary SP, Rane J. Artificial intelligence, machine learning, and deep learning for advanced business strategies: a review. Partners Universal International Innovation Journal. 2024 Jun 25;2(3):147-71.

[18] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.
[19] Rajagopal NK, Qureshi NI, Durga S, Ramirez Asis EH, Huerta Soto RM, Gupta SK, Deepak S. Future of business culture: An artificial intelligence-driven digital framework for organization decision-making process. Complexity. 2022;2022(1):7796507.
[20] Paramesha M, Rane N, Rane J. Big data analytics, artificial intelligence, machine learning, internet of things, and blockchain for enhanced business intelligence. Artificial Intelligence, Machine Learning, Internet of Things, and Blockchain for Enhanced Business Intelligence (June 6, 2024). 2024 Jun 6.
[21] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[22] Shivadekar S, Mangalagiri J, Nguyen P, Chapman D, Halem M, Gite R. An intelligent parallel distributed streaming framework for near real-time science sensors and high-resolution medical images. In50th International conference on parallel processing workshop 2021 Aug 9 (pp. 1-9).
[23] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.
[24] Pallathadka H, Ramirez-Asis EH, Loli-Poma TP, Kaliyaperumal K, Ventayen RJ, Naved M. Applications of artificial intelligence in business management, e-commerce and finance. Materials Today: Proceedings. 2023 Jan 1;80:2610-3.
[25] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..
[26] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.
[27] Oldemeyer L, Jede A, Teuteberg F. Investigation of artificial intelligence in SMEs: a systematic review of the state of the art and the main implementation challenges. Management Review Quarterly. 2025 Jun;75(2):1185-227.
[28] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.
[29] Naz H, Kashif M. Artificial intelligence and predictive marketing: an ethical framework from managers' perspective. Spanish Journal of Marketing-ESIC. 2025 Jan 2;29(1):22-45.
[30] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.
[31] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[32] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[33] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[34] Menzies J, Sabert B, Hassan R, Mensah PK. Artificial intelligence for international business: Its use, challenges, and suggestions for future research and practice. Thunderbird International Business Review. 2024 Mar;66(2):185-200.
[35] Lee J, Suh T, Roy D, Baucus M. Emerging technology and business model innovation: the case of artificial intelligence. Journal of Open Innovation: Technology, Market, and Complexity. 2019 Sep 1;5(3):44.
[36] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.

[37] Kulkov I. The role of artificial intelligence in business transformation: A case of pharmaceutical companies. Technology in Society. 2021 Aug 1;66:101629.
[38] Khan SA, Sheikh AA, Shamsi IR, Yu Z. The implications of artificial intelligence for small and medium-sized enterprises' sustainable development in the areas of blockchain technology, supply chain resilience, and closed-loop supply chains. Sustainability. 2025 Jan 4;17(1):334.
[39] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.
[40] Horani OM, Al-Adwan AS, Yaseen H, Hmoud H, Al-Rahmi WM, Alkhalifah A. The critical determinants impacting artificial intelligence adoption at the organizational level. Information Development. 2025 Sep;41(3):1055-79.
[41] Haenlein M, Huang MH, Kaplan A. Guest editorial: Business ethics in the era of artificial intelligence. Journal of Business Ethics. 2022 Jul;178(4):867-9.
[42] Goralski MA, Tan TK. Artificial intelligence and sustainable development. The International Journal of Management Education. 2020 Mar 1;18(1):100330.
[43] Gong Q, Fan D, Bartram T. Integrating artificial intelligence and human resource management: a review and future research agenda. The International Journal of Human Resource Management. 2025 Jan 2;36(1):103-41.
[44] Panda S. Scalable Artificial Intelligence Systems: Cloud-Native, Edge-AI, MLOps, and Governance for Real-World Deployment. Deep Science Publishing; 2025 Jul 28.
[45] Ghimire A, Thapa S, Jha AK, Adhikari S, Kumar A. Accelerating business growth with big data and artificial intelligence. In2020 fourth international conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) 2020 Oct 7 (pp. 441-448). IEEE.
[46] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[47] Getchell KM, Carradini S, Cardon PW, Fleischmann C, Ma H, Aritz J, Stapp J. Artificial intelligence in business communication: The changing landscape of research and teaching. Business and Professional Communication Quarterly. 2022 Mar;85(1):7-33.
[48] Doshi AR, Bell JJ, Mirzayev E, Vanneste BS. Generative artificial intelligence and evaluating strategic decisions. Strategic Management Journal. 2025 Mar;46(3):583-610.
[49] Di Vaio A, Palladino R, Hassan R, Escobar O. Artificial intelligence and business models in the sustainable development goals perspective: A systematic literature review. Journal of Business Research. 2020 Dec 1;121:283-314.
[50] Carter D. How real is the impact of artificial intelligence? The business information survey 2018. Business Information Review. 2018 Sep;35(3):99-115.
[51] Almaraz-López C, Almaraz-Menéndez F, López-Esteban C. Comparative study of the attitudes and perceptions of university students in business administration and management and in education toward artificial intelligence. Education Sciences. 2023 Jun 15;13(6):609.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 8: The U.S.–China AI Power Race: Geopolitical Dynamics, Technology, and Security Implications

1. Introduction

According to Chinese President Xi Jinping, the Chinese artificial intelligence (AI) technology is transforming production, good governance, and reshaping the world. The foundations of such imperial plans can be viewed in the development plans of AI capability by the Chinese in the organised manner of military-civil integration, a policy that guides the national resource invested in People's Liberation Army (PLA) and use of an explosive combination of leading technologies in mass surveillance. Evolving policies with these goals are generating connectivity, security, and value creation dilemmas for democratic societies, creating conditions for the emergence of an AI cold war, an intensified era of competition enveloping advanced technologies, state surveillance, economic direction, and emerging military capabilities. Understanding these driving forces is vital for effective mitigation strategies.

These dynamics shape a new phase in U.S.-China competition within a multipolar context reshaped by globalization, connectivity, and emerging technologies. Beyond economics and trade, a broader strategic framework encompasses technological conflict and a future divergence in system models dealing with society-sustaining issues. A fear of cyberwarfare, loss of technological sophistication, and perceived threats to national sovereignty and security underlie a race for control of next-generation technologies that offer new paradigms of governance and state domination. Central to these discussions is whether democracy or authoritarianism will emerge as the system model best suited to handling these matters, particularly economic direction, state surveillance, and military preparedness.

2. Historical and Strategic Context

The dynamic between the United States and China is the central theme to great-power rivalry in the 21 st century because it has resulted in a novel contest based on technological dominance between these two states. After years of intensive development in the field of artificial intelligence, the given developments are slowly transforming the military performance and have facilitated new methods of state surveillance and control. Important implications of these trends with relevance to the security and defense policy are sometimes referred to as the rise of an AI Cold War. They can not be only comprehended in the traditional meaning of military deterrence as the word implies. Rather, the policy responses should prioritize equal level to the diplomatic, economic, and the technological aspect of the AI-related competition [1,3]. This kind of framing would permit more policy action to reduce the dangers of such trends.

Scholars like Ai and Ouellet, Hu and Shambaugh, and Kellele, have pioneered the attention to AI-controlled competition between the United States and China in terms of the military competition. Such

analyses have elucidated deterrent and facilitating aspects where the government needs to intervene in various fronts, among them, multilateral collaboration framework and interoperability, technological decoupling, and supply chain resilience. The parallel line of analysis, i.e., pinpointing critical strategic questions behind the dynamics of AI, explains the current developments concerning the impact on policy decisions, risk events, and weapon technologies. Parallel to it, it illuminates the impact of AI-controlled state surveillance on privacy and civil liberties, as well as on the development of the international norms and state behavior.

2.1. The Evolution of AI Competition

It is a good idea to frame the AI competition based on a Cold War. Cold Wars do not mean an ideological issue, but another great-power struggle between competing regimes that are essentially the antagonists of the political and economic interests of each other, aiming to undermine the success of a rival. The AI allows the society to spy on and regulate people more than any other technology. The desire to have the most developed AI systems in Beijing is founded on the need of the Communist Party to ensure optimization of the surveillance state. In the United States, the application of AI, which facilitates social governance and control, is a normative issue in the area of democratic control. The government must use regulation, supervision, investment, and incentives to reduce the determination of the party regime and its allies to develop these applications.

Wanton development is not only an ethical issue; it may also lead to state failure. The desire to have such capabilities impacts on the national security factors and defines the alliances that promote the development of an AI Cold war. The result of that would eventually undermine world prosperity. Kim Gyeong-su, the chief of the Korean People's Army General Politics Bureau, delivered a speech to the East Sea Fleet sailors on July 2, 2023, emphasizing that they should stare at the enemy and never have some stupid talk..” The United States and its allies should lead the voice of common sense, invest in risk management capabilities, and strengthen cooperation with like-minded states, persuading great powers in the middle, poor, and developing groups to mitigate the emergence of an AI Cold War.

2.2. Conceptualizing the AI Cold War

The AI Arms Race has become one of the most urgent security issues in the United States, and there is growing consensus that China's concerted efforts to develop the expanded capabilities promised by an overhyped generative AI revolution require a similar realignment of strategy [4-7].

The idea of the Cold War, its emphasis on deterrence, the structure of alliances, the development of new weapons, and the statecraft in general offers a sufficient concept to evaluate the risks and counteractions that would be involved in the accelerated incorporation of AI in the PLA and other covert uses that the Chinese security services introduce. Nor do implications lie within the traditional security and military competition. The possibility of asymmetric advantage in everyday governance based on mass surveillance of the people imposed by the conglomerate of AI-enhanced digital infrastructure and the suppressive political structure provides a call to the United States and allied states to develop a consistent vision of resolving the technological, institutional, and normative aspects of the AI-enabled state.

The AI Cold War represents something different as compared to the prior arguments concerning the long-term strategic desires of the Chinese, whether Beijing desired a Sino-centric system or was a status quo power, as well as the category of great-power competition that is related to economic statecraft and sovereignty of technology. As the economy of China has been comparatively shielded by military conflict

in the Taiwan Strait, the threat of AI-facilitated instability in the United States and its allies, and in the deterrence balances, has been queued. Failure to avert disruptive developments would compound the economic damage and social costs inflicted during the COVID-19 pandemic; dramatically reduce the U.S. population's confidence in the legitimacy of democracy compared with autocracy; and reshape the international balance of power in favor of authoritarianism.

3. China's Military–Civil Fusion

The military-civil fusion is one of the key military technological innovation doctrine of the Chinese Communist Party. Its institutional and policy elements, with the creation of dual-use technology, efficient distribution of the resources of the defense industry, and building of the ecosystem of the R and D, contribute to the ability of China to take advantage of the geoeconomic competition with the United States and its allies to receive adequate resources in the domains of semiconductors and other minerals and create its capabilities in the fields of niche industries. These benefits, in their turn, lead to the proliferation and implementation of AI-enabled surveillance technology, subsequently broadening the Sino-American split in technological development in the formation of tools that can stabilize totalitarianism.. Together, these factors have fueled calls to introduce a new paradigm of AI arms control to address the risk of deterring China's military modernization and displacing it from the AI-enabled surveillance industry [7-10].

China's military–civil fusion investment model establishes a framework for military and party agencies to coordinate the allocation of funding and resources across the economy in pursuit of AI-enabled military modernization. The model enables military-industrial agencies to supplant a difficult-to-replace and financially constrained segment of the tech ecosystem responsible for cutting-edge research in niche dual-use and defense-specific technologies and to direct the economy's AI-enabled capacities toward hard-to-economically-satisfy military requirements. It thereby allows China to leverage strategic cooperation with, and geoeconomic competition in, the United States and allied economies to address risk and dependence factors that impede its capacity to create AI-optimized products and services for rolling back the U.S.-led liberal world order and for stabilizing the party's authoritarian regime.

3.1. Institutional Frameworks and Policy Drivers

China has chosen to focus on the local production of frontier technologies, particularly AI, material associated with technologies, and semiconductor chip which form the basis of military strength and national security. Such system of military-civil fusion has been created to support creation of AI-enabled surveillance and control facilities within the country, and with perspectives of broader exportation. The institutional architecture regulates the interaction between the military and civilian R&D, and manufacturing of state of art AI technologies internally [1,11-14]. The military-civil institutionalized structure involves the formal and informal defense programs. The most important non-official factors are research partnerships, which are managed by the Qinghua University Institute of Intelligence and funded by the AI 2030 project initiated by the government. Military units have in turn established their own firms, several of which are AI start-ups founded by former officers and employees. Major military end-users have also set up commercial subsidiaries to capture profitable civilian markets. The architectures of this form of state capitalism permit smarter party nuclei to use the capital of the private in the provision of military requirements and observation along with the regulation of minimal business investments, business risk, and research and advancement expenses incurred by business ventures.

The air force and army have been the first to propose military-civil fusion in AI.. The Air Force Intelligence Research Institute is focusing on China's specific value-added advantage in emerging AI-related sectors, adapting fusion technologies originally derived from military defense to economic needs, and enabling Chinese private capital to make the R&D investments required to match global leaders like Microsoft and Google [15-18].

The Army Commercial Support Fund is taking a less advanced route; its focus is to apply non-military companies as capital providers and intent on making an investment in various commerce markets that may or may not be connected to the military and national security aspects of the army..

3.2. Defense-Industrial Coordination and R&D Synergies

Recently, the Ministry of Defense of China published a White Paper named The Data-Driven Economy: A Force for the Future, which states how the data-driven economy is changing all spheres of life in China and claims that it is trying to make high-tech trade a normal activity. Apart from concerns about Taiwan, the United States has restricted trading with China in specialized chips for AI applications. These enable the rapid training of algorithms to exploit large data sets for complex social science predictions, automated surveillance, and social network sowing of political discord, general influence operations, and other capabilities. High-tech commerce is, therefore, not a one-sided proposition.

The Department of Defense's risks from China's military–civil fusion push would be lessened if the United States could strengthen semiconductor supply-chain resilience. To reduce the dependency on Taiwan, the Navy's chief of naval operations is cultivating key relationships internationally and investing to develop supplies of specialized chips in Strategic Allied countries not considered adversaries, such as Australia, the Philippines, India, and Japan [19-21]. At the same time, semiconductors are not the only functional technology required for high-end competition. Critical minerals for other high-tech areas are, therefore, essential, and countries outside of China—with the exception of the second-ranking producer—are realizing they need to cooperate on developing supplies.

3.3. International Implications and Export Controls

The use of export controls based on the need to protect national security is a two-use problem in military-civil fusion, especially in the context of the AI rivalry as a whole. Export restrictions are an essential part of the U.S. strategic rivalry with China that seeks to preemptively reduce the technological potential of the PRC through depriving it of access to state-of-the-art technological services of U.S. and other allied companies. Empirically, China's military–civil fusion strategy seeks to narrow the technological gap between the PLA and the U.S. military while exterminating the effects of U.S.–led economic statecraft. Data and information are also known to be important in the development of AI, which contributes to the emergence of interest in data-oriented military-civil fusion projects, particularly on the internal processes that focus on industry, academia, and research. In its turn, becoming overcalibrated might result in the unfortunate consequence of increasing U.S. and allied economic statecraft in a manner that will enable the empowerment of local Chinese companies and at the same time maintain access to economically and technologically vital third markets.

Empirical evidence from economic statecraft and supply chain de-risking policies indicates that the current U.S. administration is genuinely concerned over the potential military use and military–civil fusion of specific categories of emerging and foundational technologies. However, the efficacy of an expanded and more stringent U.S [22-24]. export control policy is debatable. Nonetheless, as long as the

U.S. origin technology plays a vital role in strategy and economic development of the PRC, trade flows indicate the U.S. firms are losing the battle in basic technology characteristics of supply chain resilience, competitiveness, and technology.. Chinese investments in different regions of the world—defined and shaped by distinct comparative advantages—foreshadow supportive patterns for military–civil fusion.

4. Semiconductor Dependence and Critical Minerals

he rapid development of a digital economy is finally required to rely on the availability of a sound semiconductor supply chain which is capable of supporting the present and future demand of much of AI implementation in the consumer, business, and military domains. Because the design and fabrication of advanced semiconductors require extreme expertise and specialized facilities, the risk that any nation or company would become reliant on only a small number of suppliers was acknowledged well before the emergence of AI as a powerful and widely discussed technology. The supply chain design of any silicon-based semiconductors has therefore been turning more complicated and international. However, over the last several years, there is a crisis of confidence over that architecture, both fueled by fears of the long-term U.S.-China cold war, and because of the declaration of Chinese goal of technological independence, particularly in semiconductor industry [25-28]. Such concerns were aggravated by the timing of the COVID-19 pandemic and the situation faced by Taiwan Semiconductor Manufacturing Company. Although supply chain disruptions are a recognized part of all economic activity, processes that go out of sight tend to go out of mind—until they go wrong.

Weaknesses in the semiconductor supply chain may be considered in various ways- charting its infrastructure, evaluating its capabilities locally, and monitoring international realignments. Without denying the significance of U.S. breakpoints upstream in cutting-edge manufacture, concern has focused downstream on the extent to which the United States depends on a small number of companies for the supply of a small number of products, including high-performance microprocessors. Such concerns have intensified as the ramifications of the 2022 Russo-Ukrainian War have affected supplies of semiconductor-critical neon and palladium, drawing attention to the potential for extended supply shocks caused by the geopolitical realignments. The architecture of the supply chain for critical minerals also warrants scrutiny.

4.1. Supply Chain Architecture and Vulnerabilities

China continues to depend on external sources for leading-edge semiconductor components and advanced integrated circuit design capabilities. However, the sheer volume of Chinese chips and electronic components in the world market still exerts considerable leverage for any use of export controls, no matter how ill-defined and poorly coordinated. Further, it is moving in the opposite direction of the unbalanced priorities of AI as a tool of national-security as this goes against the bleak context of investment choices, possible tendencies to purchase locally and management of cutting-edge technologies are or have traditionally been heavily business-relevant as they have always been foreign-policy-relevant. This means that most of the sophisticated technologies and, more broadly, the sub systems of complicated military and civilian systems continue to depend on perpetuated international reliance.

In 2020, China made national plans to amass some mineral resources which it considered to be the key to technical transformation and security [3,29-31]. Although the early dependence on adequate domestic production and stockpiling systems of rare earths has been changed to factor in later possibility of reliance on other important minerals to transition to a more environmentally friendly economy of renewable

sources, demand response capacities and demand maneuvers, the position of China in the production and export of rare earths has been dominant. Currently, metal stocks are declining; prices have surged across the entire period and are expected to remain elevated; potential investment decisions in new projects are shown to be complex and with substantial lead times; and completion of the underway investments is slower than anticipated.

Over the past decade, major manufacturers of electric vehicles, battery technologies, and solar panels have acknowledged the difficulties of relying on China as the main source for advanced technologies and components. As a result, alternative supply chains are being established, involving dedicated partnerships and funded by state support. Despite the corrosive influence of the US–China rivalry, leading corporations remain fully aware of the importance of a resilient, stable, and well-functioning supply chain for their business prospects [32-34].

4.2. Domestic Capabilities and Strategic Stockpiling

Out of the pre-U.S.-China trade war, semiconductor supply chain has become diversified more frequently though with frequent compromise to efficiency. Nevertheless, transatlantic and transpacific projects to re-align value chains to cover the trusted partners are on the rise, and resilience to future supply shocks is being sought towards even higher levels of economic security. The United States and the European Union have developed a world effort to mobilize investments in semiconductors and quantum technologies and Japan, the Netherlands, and the United States expounded stiffer rules and investment regulations to prevent the deviation of these semiconductors towards military applications. The establishment of new Alliance of Chip 4 between Japan, Republic of Korea, Taiwan and the United States is also concerned on the issue of supply chain security and limiting the illegal diversion of advanced chip and technology in manufacturing. However, at least in the shorter term, such localization and risk mitigation will often come at the cost of considerable trade losses and lower welfare.

China is also aiming at strengthening its local production and packaging semiconductor by making huge investments and funding to the development of new packaging terms and by pooling the demand domestically. These attempts are however being complicated by vested interests in the technology or telecommunications sectors [35-38]. Simultaneously, the American and European Union support of stockpiling critical minerals seems quite appropriate: The rapid spread of the green technologies is already creating a huge demand of such raw materials in the nearest future, where the extensive new local production capacity of the United States or the European Union (or in both cases) of the allied countries is yet not established.

4.3. Global Supply-Chain Realignments and Collaboration

The priorities of the policies proposed by the European Union, U.S., and Japan focus on the empowerment of local manufacturing capacities on semiconductor production and also globally coordinated supply chains resilience: obtaining access to vital materials, maximizing collaboration with the trusted partners, and enhancing its economic security. The current conflict in Ukraine and the increased concentration with China have solidified the attention on the links between vital mineral resources and other more economic security aspects. The statement of the summit of the Group of Seven highlights that the G7 is interested in achieving secure, sustainable and responsible supply chains of the crucial minerals and in the process of promoting a reliable access to crucial minerals by other responsible likeminded suppliers [39-41].

Lastly, the global manufacturing society is as well being influenced by the "Chips and Science Act" and the United States "Inflation Reduction Act. Both legislative documents contain a program of incentives and subsidies that focus on the manufacturing of semiconductors and electric cars, including their well-desired components and essential elements. This is the case in the enactment of the “Understanding the Increased Supply Chain Transparency of Semiconductor Manufacturing order, which represented the early attempts in ensuring that these new incentives do not simply brew an increase in the production of goods globally but rather that true production capacity is created in the country.

5. AI-Enabled State Surveillance and Democratic Constraints

Important dimensions of the AI power race can be understood through the extent to which these technologies enhance trends towards state-level control, surveillance, and oppression. These developments may create imbalances not in favor of the United States and its allies. While there are numerous technological capabilities and applications—some, but not all, military-based—that are critical to state design and implementation of the kind of intelligent governance envisaged by the Chinese leadership, these are not the focus of this section. Rather, the analysis centers upon the application of AI technology to domestic governance and control, specifically state surveillance and the suppression of dissent. Such applications are already widely resource-intensive and enhanced use can be expected as a consequence of technological development.

The artificial intelligence will become a very important element of shaping the surveillance system that will be required by the political authority to establish the system of political control and repression of the dissent [42-45]. One of the symptoms of the so-called AI-enabled surveillance race is a rivalry over creating AI that is applied to the state level to govern populations and territories instead of assault or defend against opponents involved in the military conflict.. The race is witnessed largely through the lens of China as the leader in applying these technologies for governance and control. The authoritarian nature of the state makes examining the negative effects of such capabilities and applications easier than it would be in the case of a democracy where privacy, civil liberties, and the protection of dissent may be viewed as priorities.

5.1. Technological Tools for Governance and Control

China’s growing capacity for state surveillance is an alarming development that implies profound implications. The rapid advances in AI technology augmenting state surveillance capabilities and the growing convergence between authoritarian and democratic nations which regards surveillance as a technological tool for maintaining social and national stability have been noted in detail. States’ governance and regulation of AI by dictating the value orientation and scale boundaries of relevant technologies is inherently a service orientation, a limiting effect on technological development that needs to be tested through specific cases. The maintenance of AI privacy protection that seeks to benefit civil society must be combined with the promotion of mainstream values in the market economy.

During the Cold War, the United States and the Soviet Union were ideologically confronted to some extent, but the two superpowers generally accepted liberal or relatively open ideologies. Private industry was the main supplier of military technology, while the military monopolized official counterintelligence and intelligence functions. However, state-military relationships in China’s industrial and technological development differ from those of the United States and the Soviet Union. The government has established a system of military-civil fusion that encourages private companies to develop new technologies. Such

technologies are not limited to military applications and can be used to establish new industrial systems that are more competitive than the old SSC model. At the same time, however, the Chinese government places a specific technology direction on the development of private companies. In the intelligent age, the advancement of science and technology has outperformed the retrogression of human civilization.

5.2. Comparative Governance: Surveillance Architectures

The Chinese state has deployed AI-based systems to enhance its governance capacity across diverse domains, including public safety, urban management, environmental protection, and health care. These efforts have combined significant resources, political commitment, advanced technology, and sensitive data-gathering capabilities to develop an exquisite state-surveillance architecture containing many tools of state control. China has established the capacity to monitor the activities of its domestic population at an unprecedented level of detail—one that would likely generate an international outcry were it to be deployed against the United States or other democracies. Continued U.S. dependence on Chinese production of critical minerals could enable China to slow or stop such supplies to the United States. That policy tool would likely appeal to the Chinese Communist Party at a time of economic weakness, increasing the domestic political salience of this decision while limiting the global economic costs. In this regard, being able to gain such supplies without going through the United States would confer a significant (if asymmetric) advantage.

While the capability to deploy these approaches within China provides an expansive governance toolbox for the Chinese Communist Party's increasingly repressive regime, there are serious international implications [46-48]. Many of China's AI-based systems for surveillance and social control within China would violate international norms of privacy and civil liberties if deployed in the United States, European Union, or other democratic societies. Chinese leaders appear willing to export components of their surveillance-state architecture that depart from such norms, and relative demand for these exports is expected to grow in many countries. Balancing privacy and civil liberties with broader societal goals is a complex and ongoing challenge within many democracies. Policies that reduce domestic demand for AI-based state-surveillance and social-control technologies below that of other democracies can help reduce supply pressures on other countries.

5.3. Privacy, Civil Liberties, and International Norms

The novel instruments of technological governance that the Chinese Communist Party (CCP) has put in place provide insight into the state's priorities, especially regarding the resolution of tensions between control and efficiency. They also illustrate possible relations between state surveillance capacity and regime type. However, because the technological tools for governance and control do not exist in isolation, it is equally important to understand how they interact with the state's material, regulatory, and moral bases of governance; the supply chains behind the cameras, facial-recognition AI systems, the storage and analysis of enriched datasets, and cloud computing; the evolution of the state's role in the economy; the evolution of the architecture of manifesto; as well as alliances, support networks, and supply chains in a world out of sync.

The fast-accumulating technologies of the twentieth-first century—AI, biotechnology, quantum computing, second-generation Internet/5G, brain-computer interface, and the Internet and intelligent systems of things—support cold-war scenarios of AIs acting on behalf of governments. But for a given country, the risk group has continuously emerged as a coalition of various ethnic and religious minorities.

For A. Shcherbenok, the cold war, similarly to Albert Einstein's view of the human race, can be publicly inexplicit, formulating a Middle East Ku-Klux-Clan movement in the Islamist countries of the Middle East. What it may not seem clear is whether all of these historical memories going thousands of years back, and the economic aspirations and the quest of modern society for the happiness and serenity of life for everyone, could lead to using AI for life without hesitation, with the world in harmony and peace.

6. Strategic Implications for Security and Defense

AI and machine learning dramatically lower the cost of deterrence, enhancing the credibility of limited threats. The risk of unintentional escalation in a conflict between the United States and China may consequently vary considerably in different scenarios. Ex–U.S. defense officials are concerned about the risk that China may attempt to alter the status quo in the Taiwan Strait through limited force.

Deterrence challenges extend beyond Taiwan. AI currently enhances both powers' military operations but may not do so equally [4,49]. AI may enable China to use coercive threats to deter allied intervention by dramatically increasing the costs of intervention without jeopardizing China's vital interests and territory. New warfare concepts in China's military–strategic thinking and the establishment of joint operational headquarters are consistent with this growing capability. More broadly, success in harnessing AI for military operations in an armed confrontation with a peer could determine the outcome in theatres of armed conflict worldwide.

The fact that there has now been the development of AI-enabled weapons systems to be utilized in crises and conflict only adds to the complications of the issue of escalation control. The AI-controlled weapons will be deprived of fully developed protection, and overcoming the unwanted escalation will be hard, and it will have an argument in favor of deterrence and strategic stability. Such concerns are enhanced by the recent events in history, especially the Ukrainian conflict and the level of collaboration between Russia and China..

6.1. Deterrence, Alliance Architecture, and Interoperability

AI Cold War Dynamics Continued rivalry in high-technology military systems - AI-powered capabilities that cramp the boundary between peace and conflict, and AI-ML research and development domains, architectures, and ecosystems - are all indicative of geopolitical abysses. Dueling deterrence strategies of the leading AI powers, their visions of alliance architecture, stable great-power military relations, and the creation or avoidance of conditions that might lead to inadvertent war determine the nature of great-power competition. AI-enabled military capabilities can now advance, diminish, deter, or defeat vital interests. Coalition warfare becomes the default response to aggression; increasingly, future war is conceived of as a coalition encounter at global or near-global scale. The advance of machine learning and related technologies may even make nuclear war thinkable, not least during the transition of US—China relations from competition to conflict. Inevitably, the alliance and interoperability architectures required to conduct combined coalition warfare at great scale and distance shape national priorities and investments in advanced military technologies.

Democratic countries with broad access to advanced military technologies, and capable of operating them effectively in the context of a combined oalition military campaign, are in a strong and advantageous position. Such countries can make credible deterrence and assurance commitments through both capability development and declaratory policy. They are able to create a coalition form that would attract the largest military players nearest to the conflict area and hence enhance the size and efficiency of their

deterrence. Stability and security of those participating in the deterrence connection enhancement is enhanced as increases of deterrence capacity happen as the connections of these display postures cascade or cross-flow, offering better stability and security. During military confrontation, lack of interoperability and access to sophisticated military skills will be a major weakness. At the point where common operating procedures, information-sharing protocols, and language capabilities are hardwired, the benefits of the AI race are mitigated or canceled, and the discouraging effect lessened.

6.2. Emerging Military Technologies and Risk Scenarios

Developments in AI and associated ML technologies will change the types of military operations that states can conduct. Cyber operations can achieve a greater degree of stealth, precision, or deniability owing to assisted malware design and the use of natural-language processing. The conduct of disinformation campaigns can likewise benefit from assisted message creation focused on particular audiences. Offensive drone swarms may be adapted to new use cases, particularly in relation to rapid decision-taking. In these and other areas such as advancing capabilities for tracking ISR, oil-spill detection, or identifying bomb-making facilities, AI can enable advances that lower the risk of detection, increase the probability of success, and present political benefits through enhanced deniability.

At the same time, other military technologies are being produced using non-AI technology that could be vulnerable to advanced AI capabilities. Information-gathering and reconnaissance operations, together with the provision of supporting fires, could be rendered subject to a higher level of danger through the development of sub-strategic offensive strike complexes and unmarked drone strikes through third-party proxys [1,3,5]. The ability of major armed forces to accurately apply economic strength against smaller adversaries with lower economies of scale is being rendered more difficult by the progressive development of longer-range precision systems that support a threat to critical infrastructure in those states. The development of AI in military applications can therefore provide both new opportunities and new risks, and emerging military technologies should also be considered in terms of deterrent effect and military–political responses.

6.3. Economic Statecraft and Technological Sovereignty

As competition over AI capabilities intensifies, increased state intervention to bolster domestic industry in the United States, like China's economic statecraft, is to be expected. National AI strategies like the U.S.–Japan 5G accord, the European Chips Act, and the Chips and Science Act are government responses aimed at achieving regional technological sovereignty. The United States government has imposed sweeping controls on semiconductor manufacture in China, prohibiting exports of cutting-edge equipment and U.S.-made chips for AI and supercomputer applications in Chinese firms. Such decisions seek to limit China's access to the advanced computing capabilities that can support AI, machine learning, and deep learning applications. U.S. measures also curb investments in the Chinese technology sector and impose restrictions on AI-related technologies, via the U.S. Bureau of Industry and Security.

While experts contest the sufficiency of these initiatives for technological decoupling from China, some regional states are establishing a framework for the development of a partner ecosystem and proximity supply chains for sensitive technologies. At the same time returning to fiscal, monetary and trade policies based on the priorities of national interests and the long-term capabilities of the industries are aimed at securing supply chains of the most necessary items and ensuring that domestic industries, which have sensitive technologies, are not dependent on foreign countries. The future regulation of such products at

the state level is explicitly aimed at establishing and preserving the technological sovereignty. These measures are opposite to regulatory approach aimed at promotion of free trade and fair competition..

7. The Emergence of an AI Cold War: Theoretical and Practical Implications

On top of these lessons, the idea of an AI Cold War between the United States and China can be used to explain how the larger trend of great-power rivalry, as well as the U.S. and its allies policy, can be understood in this categorization. The studies of U.S.-China rivalry, along with its consequences to the global order, can appear as being out of touch with the pertinent theory. Departing from the historical Cold War, popular approaches—certainly intuitive and laden with resonance—have focused instead on Leninist ideology and governance model, differences in regime-type, trade, and industrial policy relative to globalization, and thus on economic statecraft and tech trade conflicts. But in a theoretical landscape anchored by the concept of two-dimensional rivalry, the political-economy dimension is exposed as an imperfectly substitutive treat.

Framing military, technological, and economic competitiveness differently leads to very different action-guidance, and highlights distinct areas of concern. Of course, the many previous great-power competitions incorporating some form of trade competition, including the Cold War, remain relevant; rupture of international trade in semiconductors, microelectronics, and AI training systems could hinder, although not completely prevent, the development of military capabilities. Nonetheless, the effort to retool a broader set of trade relations, especially with Japan and the EU, appears less a first-order objective than a desideratum for ensuring that deterrence remains robust. In contrast, the two-dimensional rivalry framework elevates the two great military technologies that have historically triggered high risks of armed conflict, namely advanced military capabilities and integration of military alliances, to prime focus.

7.1. Debates and Paradigms in Great-Power Competition

Beijing's transformational strategy for AI, while it may be understood primarily as an all-encompassing effort to advance the state's military modernization and economic development, entails an expansion of the power and operational scope of the Chinese Communist Party (CCP). For both countries, the race is not merely about developing AI technologies; it is also about state governance specifically enabled and supported by AI technologies and derived tools. As such, scholars and policymakers should consider whether an "AI–surveillance" Cold War is developing within the broader context of great-power competition [7,50]. Recent events have accentuated the state of competition and the risks of confrontation between liberal democratic and authoritarian states, raising the specter of global technoauthoritarianism culminating in a digital iron curtain of bifurcating world orders.

In general, however, great-power competition remains poorly theorized. The AI-surveillance-race brings the shortcomings of the current debates on theories of victory, and on the implication of economic statecraft that aimed at technolo!gical decoupling and supply-chain resilience. Helen Milner offers a more beneficial theoretical perspective in terms of tit-for-tat types of behaviors that react and evolve to the actions of the competitor. This view implies that the United States and allies ought to accentuate the impacts of the surveillance technology on governments, personal liberties, and human dignity, and aim at controlling the competition with a par!ticular focus on escalation danger and coalition administration. More detailed options in policies affect the creation of AI governance models to ensure responsible innovation, international security sanction regimes in particular with reference to AI-driven surveillance

and repression, and diplomacy with the creation of AI technological standards that enable market choice and address related security risks. Last but not least, to address the issues of the AI Cold War on the stability and world governance, the United States and its allies may make efforts to avoid the development of enmity alliances in the sphere of AI..

7.2.Policy Options for the United States and Allies

Despite a Realist assumption of self-interest driving the intensification of U.S.–China competition, a variety of explanations and a broader set of concepts have been applied to define the nature of the emerging conflict. These include not just the increasingly formal recognition of ideology as a significant driver of great-power conflict, but also the argument that technocratic governance, the diffusion of increasingly powerful technologies, the circulation of people and ideas, and the presence or absence of nuclear deterrent relationships shape the understandings and preferences of authorities in Beijing and Washington about how and why to compete with one another . In contrast to the events and changes of the past three decades—characterized by the growth of highly integrated multilateral supply chains and deep transnational economic linkages underpinning both globalization and Grotrian and liberal hopes for a long-term Hegelian movement toward a China that is both free and wise—the future is increasingly understood in terms of zones of heightened rivalry, reduced economic and technological interdependence, and greatly increased risk of accidental or inadvertent conflict.

Staying at par with the Chinese aggressiveness in the AI industry is complex in terms of opportunities, traps, and challenges to the United States and its allies. They need to redesign initiatives, which will build upon the power of collaboration, faster supply chain, and enhanced priority to the combination of ambition and prudence, risk of awareness and technology, with minimum fragmentation. In order to realize the innovative potential of generative AI and use hundreds of ways it can be corrupted and targeted at evil purposes, significant investments in the regulation are necessary- and not only in the costumes of Halloween spiders and bees. Leaders also face two greater challenges: how to deter China's long-term ambitions from achieving major strategic and quantitative advantage, even in the context of states-blind politics.

7.3. Risk Management, Alliances, and International Collaboration

The emergence of an AI Cold War between the United States and China is a rapid process that requires careful monitoring and ideally proactive management. Internally, the primary focus should be on developing complementary capabilities. Although AI-enabled technological builds represent an opportunity for deterring actions through the deployment of new tools, the balance is also highly unstable and early-stage techniques and operational concepts remain fraught with danger. Missing incentives may dissuade actors from setting in motion even the most important builds, such as the munitions industrial base. The one real area for temptation is within Cyber and generally within the Information Domain, where early developments carry the combined virtues of false evidentiary proof of capability and little risk of immediate exposure.

Within the Allied framework, the United States must steer NATO–AI agendas and the Alliance's overall strategic messaging towards tempering the natural disinclination of the European members to overcome their historical wariness for too close an alliance with Washington's Polar–Asia strategy. The allure of an economic decoupling in the trade route running to the East Coast is a potent temptation but doing so would leave the United States with the problem of both balancing and deterring China. with the help of

International Security Cooperation Agreements, not least in Chapter V of the charter of NATO, collaboration with the Polar-Asia players can be strengthened, though only as long as endeavours are mandatory in Areas and Lines of these poles and are mindful of the politics, economics, and effects of said so.. The AI-enabled technological builds of like-minded actors outside the Alliance–Geo group are likely to remain much closer to the broader principles of Democracy and their associated Civil and Political Rights.

8. Implications for Policy and Strategy

Allied technocratic governance and regulation of AI become that much more pressing in the context of these dynamics. The exacerbation of a power competition centered on the applicability of AI reveals risk mitigation—through both policy enabling critical U.S. or allied capability and responding to the risk of dual-use technology transfer—as critical. Nuclear deterrence considerations, alliance due diligence ensuring mutual cohesion, and attempts to contain subsequent AI-enabled malicious operations drive U.S. policy and narrative and support proposals for building functional firewall structures to preserve vital civil liberties and protect against digital colonialism. The focus remains, however, primarily on establishing an impact-oriented ethical or values-based model of AI governance.

Such an approach to AI regulation remains at risk where governing agents incapable of effective governance maintain access to the full spectrum of AI tools. Like the technology itself, governance requires a continual balancing of opportunity against risk. Technology transfer between great powers remains fundamental to the sustainability of the mutually reinforcing standards paradigm and preeminently placed by the Chinese Communist Party in an AI warfo-rm—Modern China's AI Cold War. This has accrued much criticism, where it has been learned to institutionalize a more lasting, capability-based approach towards international AI rules, without absence of portrayal of some Maginot Line, which is merely the postponement of the compulsory expansion of such tools and technologies into approaches of governing the populace that they desire.

It is unclear to what extent the leaders adhere to such lessons, and it is becoming more prone to manipulation by the third countries. But there is strain between Western technological giants and crypto-businessmen, getting attracted to the business prospects and an unhindered access to the latest arsenal of AI gadgets.. Like the capital markets that support them, the availability of leading-edge applications currently absent in non-liberal jurisdictions threatens to vitiate the founding proposition of a values-centered approach to AI—democracy. Hence, the risk of global collapse and failure to contain the use of AI to support ordoliberal political ideologies are prompting the establishment of functional variants of "modern technology control" machinery, which have undergone substantial evolution since the existing regime was founded.

8.1. Recommendations for Governance, Regulation, and Oversight

As global competition accelerates, there are five significant areas that require careful governance, regulation, and oversight. First, establishing agreements on competition policy involving the leading AI powers is needed. Second, it is essential to come up with norms and guardrails to AI applications in the military sphere. Third, it would be more essential to restrict the application of generative AI to generating strongly persuasive fake content. Fourth, designed mechanisms and procedures ought to be established to minimize the threat of escalation of the accidental AI-enabled error of calculation. Fifth, the threat of

injury with the ill intent in malicious utilization of AI-permitted falsehood on political and economic marketplaces needs to be considered.

White-collar governments across the world have a decision of regulating AI development in a responsible manner especially with the strategic rivalry between the United States and China. The high pace of advancement of AI technologies increases the challenge in creating proper guardrails since the situation in the industry today can be described as self-regulatory. The trend of state control or protectionism will continue to rise with views of the political leaders of both countries already putting technological advances in the context of national security.

8.2. International normative frameworks and diplomacy

Efforts to devise government guidelines for trustworthy AI and other responsible technologies must take into consideration the challenges posed both by emerging AI weapon systems and AI-enabled state surveillance. The increasing trend of nondemocratic states towards complete affinity of weapons and their military-civil convergence of AI, face recognition, and expensive data manipulation instruments has highlighted how this could be utilized to further government ends, and certain threats it poses to human individuals and the ideals of freedom in open societies. In the most acute part of any international nondiscrimination-normative framework governing the future of AI, there must be guaranteed the protection against direct application by AI technologies against civilians at a national and even global level..

Beyond these imperative efforts to construct defenses against potential unwanted outputs of AI, parallel initiatives are required that seek to adapt international relations and established diplomacy to the already changed international environment and its new technologies. The complexity and increasing conflictuality of the relations between the United States and its major allies and partners on one side and China and Russia on the other side require more active preparations, remedies, and escapes than have so far been attempted. These preparations must aim at risk containment and management, from the bottom up and the top down.

9. Conclusion

This volume has considered the nature of the geopolitical competition from a technological perspective, particularly in artificial intelligence. At the same time, it is also examined whether the present stage of great-power rivalry can be thought of as an instance of cold war perceived as extreme polarization of two blocs on the basis of their rivalry in terms of advanced technology, geopolitical dominance, securing and maintaining the world order one that represents their underlying value system, either liberal-democratic or authoritarian. It has explored how political economic and military trends related to technology reflect that competition, on key dimensions in the AI field. The summary comes to the conclusion that the AI-enabled stage of great-power intensification is a more radical variant of authoritarianism supposed to question democratic standards and supporting principles.

The study has conceptualized the emergence of an AI cold war: the prospect of militarized technical–scientific bifurcation; the competing supply chains of critical inputs such as semiconductors; the application of advanced surveillance technologies to societal governance; and their consequent impact on the foundations of international relations and global peace. In addition to supporting technologies required to sustain its own interoperability and deterrence, it finds that the United States and democracies are

compelled to engage in risk mitigation through economic statecraft and preventive measures designed to inhibit the use of AI for totalitarian state repressio

References

[1] Abrokwah-Larbi K, Awuku-Larbi Y. The impact of artificial intelligence in marketing on the performance of business organizations: evidence from SMEs in an emerging economy. Journal of Entrepreneurship in Emerging Economies. 2024 Jun 13;16(4):1090-117.

[2] Agarwal P, Swami S, Malhotra SK. Artificial intelligence adoption in the post COVID-19 new-normal and role of smart technologies in transforming business: a review. Journal of Science and Technology Policy Management. 2024 Apr 18;15(3):506-29.

[3] Mohapatra PS. Artificial intelligence and machine learning for test engineers: concepts in software quality assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:17.

[4] Ahmed AA, Agarwal S, Kurniawan IG, Anantadjaya SP, Krishnan C. Business boosting through sentiment analysis using Artificial Intelligence approach. International Journal of System Assurance Engineering and Management. 2022 Mar;13(Suppl 1):699-709.

[5] Alawadhi SA, Zowayed A, Abdulla H, Khder MA, Ali BJ. Impact of artificial intelligence on information security in business. In2022 ASU International Conference in Emerging Technologies for Sustainability and Intelligent Systems (ICETSIS) 2022 Jun 22 (pp. 437-442). IEEE.

[6] Beheshti A, Yang J, Sheng QZ, Benatallah B, Casati F, Dustdar S, Nezhad HR, Zhang X, Xue S. ProcessGPT: transforming business process management with generative artificial intelligence. In2023 IEEE international conference on web services (ICWS) 2023 Jul 2 (pp. 731-739). IEEE.

[7] Chen J, Lim CP, Tan KH, Govindan K, Kumar A. Artificial intelligence-based human-centric decision support framework: an application to predictive maintenance in asset management under pandemic environments. Annals of Operations Research. 2025 Jul;350(2):493-516.

[8] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.

[9] Chowdhury S, Budhwar P, Wood G. Generative artificial intelligence in business: towards a strategic human resource management framework. British Journal of Management. 2024 Oct;35(4):1680-91.

[10] Davenport TH, Ronanki R. Artificial intelligence for the real world. Harvard business review. 2018 Jan 1;96(1):108-16.

[11] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.

[12] Kanbach DK, Heiduk L, Blueher G, Schreiter M, Lahmann A. The GenAI is out of the bottle: generative artificial intelligence from a business model innovation perspective. Review of Managerial Science. 2024 Apr;18(4):1189-220.

[13] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.

[14] Kar AK, Kushwaha AK. Facilitators and barriers of artificial intelligence adoption in business–insights from opinions using big data analytics. Information Systems Frontiers. 2023 Aug;25(4):1351-74.

[15] Kerzel U. Enterprise AI Canvas Integrating artificial intelligence into business. Applied Artificial Intelligence. 2021 Jan 2;35(1):1-2.

[16] Kulkov I. Next-generation business models for artificial intelligence start-ups in the healthcare industry. International Journal of Entrepreneurial Behavior & Research. 2023 May 4;29(4):860-85.

[17] Kumar S, Lim WM, Sivarajah U, Kaur J. Artificial intelligence and blockchain integration in business: trends from a bibliometric-content analysis. Information systems frontiers. 2023 Apr;25(2):871-96.
[18] Maslak OI, Maslak MV, Grishko NY, Hlazunova OO, Pererva PG, Yakovenko YY. Artificial intelligence as a key driver of business operations transformation in the conditions of the digital economy. In2021 IEEE International Conference on Modern Electrical and Energy Systems (MEES) 2021 Sep 21 (pp. 1-5). IEEE.
[19] Mishra AN, Pani AK. Business value appropriation roadmap for artificial intelligence. VINE Journal of Information and Knowledge Management Systems. 2021 May 31;51(3):353-68.
[20] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[21] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[22] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.
[23] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[24] Rana NP, Chatterjee S, Dwivedi YK, Akter S. Understanding dark side of artificial intelligence (AI) integrated business analytics: assessing firm's operational inefficiency and competitiveness. European Journal of Information Systems. 2022 May 4;31(3):364-87.
[25] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.
[26] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..
[27] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[28] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.
[29] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.
[30] Reier Forradellas RF, Garay Gallastegui LM. Digital transformation and artificial intelligence applied to business: Legal regulations, economic impact and perspective. Laws. 2021 Aug 27;10(3):70.
[31] Sandeep SR, Ahamad S, Saxena D, Srivastava K, Jaiswal S, Bora A. To understand the relationship between Machine learning and Artificial intelligence in large and diversified business organisations. Materials Today: Proceedings. 2022 Jan 1;56:2082-6.
[32] Schneider J, Abraham R, Meske C, Vom Brocke J. Artificial intelligence governance for businesses. Information Systems Management. 2023 Jul 3;40(3):229-49.
[33] Panda S. Scalable Artificial Intelligence Systems: Cloud-Native, Edge-AI, MLOps, and Governance for Real-World Deployment. Deep Science Publishing; 2025 Jul 28.
[34] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.
[35] Shwedeh F, Alzoubi HM. Creating and Evaluating Instructional Java Programming Codes with Utilization of Artificial Intelligence for Customized Business Requirements. InInternational Scientific Conference Management and Engineering 2024 Jun 23 (pp. 281-286). Cham: Springer Nature Switzerland.

[36] Singh S, Goyal MK. Enhancing climate resilience in businesses: the role of artificial intelligence. Journal of Cleaner Production. 2023 Sep 15;418:138228.
[37] Sjödin D, Parida V, Kohtamäki M. Artificial intelligence enabling circular business model innovation in digital servitization: Conceptualizing dynamic capabilities, AI capacities, business models and effects. Technological Forecasting and Social Change. 2023 Dec 1;197:122903.
[38] Storey VC, Yue WT, Zhao JL, Lukyanenko R. Generative artificial intelligence: Evolving technology, growing societal impact, and opportunities for information systems research. Information Systems Frontiers. 2025 Feb 25:1-22.
[39] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.
[40] Svetlana N, Anna N, Svetlana M, Tatiana G, Olga M. Artificial intelligence as a driver of business process transformation. Procedia Computer Science. 2022 Jan 1;213:276-84.
[41] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.
[42] Toniolo K, Masiero E, Massaro M, Bagnoli C. Sustainable business models and artificial intelligence: Opportunities and challenges. Knowledge, people, and digital transformation: Approaches for a sustainable future. 2020 Apr 23:103-17.
[43] Vardarlier P, Zafer C. Use of artificial intelligence as business strategy in recruitment process and social perspective. InDigital business strategies in blockchain ecosystems: Transformational design and future of global business 2019 Nov 10 (pp. 355-373). Cham: Springer International Publishing.
[44] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.
[45] Verma C, Vijayalakshmi P, Chaturvedi N, Umesh U, Rai A, Ahmad AY. Artificial Intelligence in Marketing Management: Enhancing Customer Engagement and Personalization. In2025 International Conference on Pervasive Computational Technologies (ICPCT) 2025 Feb 8 (pp. 397-401). IEEE.
[46] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.
[47] Wamba-Taguimdje SL, Fosso Wamba S, Kala Kamdjoug JR, Tchatchouang Wanko CE. Influence of artificial intelligence (AI) on firm performance: the business value of AI-based transformation projects. Business process management journal. 2020 Nov 2;26(7):1893-924.
[48] Xu JJ, Babaian T. Artificial intelligence in business curriculum: The pedagogy and learning outcomes. The International Journal of Management Education. 2021 Nov 1;19(3):100550.
[49] Shivadekar S, Mangalagiri J, Nguyen P, Chapman D, Halem M, Gite R. An intelligent parallel distributed streaming framework for near real-time science sensors and high-resolution medical images. In50th International conference on parallel processing workshop 2021 Aug 9 (pp. 1-9).
[50] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 9: Securing America's Critical Supply Chains with AI: Minerals, Manufacturing, and Energy

1. Introduction

Since the global economy is becoming interconnected it denotes the ability to access necessities depends on an extensive range of intelligent systems, such as sensors, analytics, machine learning, industrial control, logistics, and cyber defense. The presence and services of such systems, in its turn, can be affected by various threats, including the supply chain disruptions, production, logistics, and distribution errors, as well as cyberattacks, which have the potential to take place at any part of the chain and spread to the remotest of locations. Command of artificial intelligence (AI) to enhance supply chain security has also received a lot of focus. The research community has however been much involved with the mitigation of risks especially to the automotive sector and semiconductor and much less on the solutions that are to be utilized to mitigate critical supply chains. The chapter opens up the opening based on AI-enabled supply chain security by investigating 3 most important, and potentially vulnerabilities, types with associated research challenges: the mining products and extraction process, the global semiconductor ecosystem, and the resilience of the energy power grid disruption disrupting the cyber-physical infrastructure of the modern economy.

The greatest concentration of intelligent systems, and therefore the most direct pathways to support supply chain risk mitigation and management, is in product demand forecasting and generation. Demand forecasting stimulates connected services in weather prediction, supply chain and logistics optimization, Link prediction, and other support operations, and it drives the use of AI to mitigate cyber-physical security risks. Advanced Metering Infrastructure (AMI) continuously monitors and regulates the global electricity grid with low-cost sensors which produce stream after stream of time-series data, which offers analytics and machine-learning opportunities to predict, avoid, and reduce disturbances spreading through the grid and to optimize the performance of federated operations. Grid connected batteries facilitate the Load Shifting and Demand Response which offer predictive load forecasting. AI, in particular, generative and predictive models will provide additional comprehension of grid functions by forecasting accurate load demands with demand-response beyond battery storage, controller of distributed resources in model-predictive, and camera systems purchased on clouds to guarantee enhanced security in cyber physical. Interdependencies in the supply chains that are reliant on the availability and quality of electricity, and mitigation measures against cyber and natural and accidental threat to the availability of electricity and quality need to be considered in supporting the resilient economy.

2. AI-Driven Monitoring of Rare-Earths and Critical Minerals

To inform and support the formulation of state and federal supply-chain policies that protect U.S. economic, social, and political resilience and security, AI technologies can be employed to identify origin

and route of critical minerals throughout their life cycle. Relying on the concentrated supplies of the important minerals, and most importantly, the rare earths, has both the political, economical and military consequences. Ways of allowing AI-controlled tracking of rare-earth and other valuable minerals are welcome. The best possible methods would allow tracking continuously and in real time; leveraging on easily accessible third party data; reducing the need to have specialized training data; robust performance across supply chain legs; real time assessment of supply and demand risk; and risk allocation of the ethical, legal, and environmental risks across supply [1,2].

Three AIs enabled in monitoring critical minerals through their supply chain are discussed. Systems of country-level tracking of rare-earths and these critical minerals are set; a semi-automated system of supply-chain transparency mapping is created; and a locus area is suggested towards the evaluation of ethical, legal and environmental issues relating to mineral extraction and refining..

2.1. Tracking methodologies and data provenance

Supervised and unsupervised solutions based on novel graphical methods, socially relevant scrutiny, and a versatile framework for data collection ground the authenticity, availability, and completeness of the information used to develop the global mineral supply-indexes. The underlying principles of scientific inquiry serve as a litmus for data quality, associated methodologies, and specialized knowledge: novelty, clarity, accuracy, precision, relevance, variety, and reliability. The historical, prospective, and predictive characteristics of the monitored phenomena are determined by the scientific consensus on the basic fact.

The study of the rare-earth elements (REE) trade has been fundamentally descriptive, as systematic collections of data on REE flows and stocks have been reported only recently. Nowadays, the availability of data on mineral supply chains is abundant, but the extent and integrity of the repositories are critical and urgent questions, because international development cooperation and various other initiatives to secure future supplies usually make explicit or implicit assumptions about the reliability of the repositories [3-5]. Although repositories house much relevant data, subjects such as the politics of climate change or critical minerals for green energy production require scrutiny beyond the databases concurrently operated by Finland's Geological Survey and the United States Geological Survey. While comprehensiveness is thus impractical, the different types of description and inquiry that the communities of scientists have addressed can be systematized within the area of supply-chain transparency.

2.2. Supply chain transparency and risk assessment

Disruption of the global supply of minerals required for defense applications and many more commercial products can have serious repercussions across vital sectors of the economy. Specifically deploying AI to enhance transparency across mineral supply chains, particularly for rare earths but for others as well, would help to mitigate risks of abrupt supply reductions, including from geopolitical shocks. Such approaches would also permit identification of sources of pollution and unsatisfactory labor practices.

Nonemerging stability of rare-earth and critical mineral supply is crucial given the centrality of those materials in the production of a wide range of economic sectors and products. A potential countermeasure in the substitutes of the rare-earths, in devices, such as demand, was another potential mitigation strategy, however, the effect of actual substitutes will not be significant, and a Canadian-owned grower, Troilus Metal Corporation, said that, "the ability of the domestic industry to absorb such shock is limited Director of Foundry The 2022 semiconductor plan by Vice President Kamala Harris highlighted the need to create a more discreet supply chain and inventories of the technology both military and civilian applications so

vital in the daily lives of people. Hurricanes and droughts are likely to continue being a persistent threat to the home-based production, but Taiwanese manufactures approximately 30 percent of imported semiconductor packages. Interruption with the delivery of the supplies of minerals, or metals to military defense projects may therefore cause an extreme slowing of readiness and recovery operations [6,7].

Continued manufacturing of semiconductors and their components as well as rare metals is key to supporting advanced technologies driving the growth of industries such as aerospace, transportation, and power. Tracking and exploring alternatives for crucial materials suppliers, therefore, will also enhance detective and deterrent strategies of national service and government organizations. Policy measures to guarantee supply chain security involve assessment of the flow from resources to products in the market. Recent work on supply chain transparency for critical materials provided models to facilitate an innovative search, design, and monitoring of the supply of rare elements.

2.3. Ethical, legal, and environmental considerations

Mining and refining have important ethical, legal, and environmental aspects beyond cybersecurity, yet information bottlenecks have stymied these discussions. From an ethical standpoint, market-distorting procurement policies that fuel violations of human rights, labor rights, and environmental protection can be challenged only through better supply chain transparency. The stringent enforcement of the current standards entails the need of reliable data, and Assurance x Validation x Trust x Action can be used as a guiding tool, so the lack of traceability data on the minerals and critical substances sold to the United States must be made up. Because of the way ChatGPT has recently become a full-fledged Generative Pre-trained Transformer, a prototype of the same should be developed in the case of the rare-earths and critical minerals supply chain.

The scrutiny by the law can be narrower. Other than fighting the trade in conflict minerals, there is also current laws against sourcing of minerals at unregulated, unmonitored, or unsafe production, mining as well as refining areas within Canada, Central Africa or West Africa, or from child sourcing or forced labor. Moreover, the tariff act of the year 1930 permits the customs officers to search, discriminate goods that are unfairly traded or illegal within the United States, and levy more duties, taxes, or other limitations, as required. Nevertheless, the exercise of the existing laws fully is subject to the condition of the existence of convincing provenance information.Supply interest is insufficient to deter social and environmental abuses in the production of rare-earths and critical minerals. These abuses are difficult to substantiate in the absence of transparent supply chain information that supports public sector investigative and enforcement efforts.

Environmental considerations are also pressing. Mining at nickel's frost line risks irreparable human and ecological damage, as plastic pollution illustrated so vividly; it must therefore be disallowed without exception. Experience with COVID-19 vaccines replicating viralytics and AstraZeneca's global follow-up to the pollinator extinction-promoting neonicotinoid class of insecticides emphasizes how, without trustworthy supply chain monitoring, proven technology and production locations will not be rescaled when and where needed, opening the door to exponential--or worse--collapse.

3. Cybersecurity Risks in Mining and Refining: An UAMY-Relevant Perspective

Metallic and mineral supply chains have become a heightened security concern due to the accelerating trend of cyberattacks on mining operations and metal-processing facilities [2,8-10]. While national security-related attention has centered on behalf of specific metals and minerals, the threats to the entire

mining and refining sector must be considered. These attacks are likely to increase over time due to the interplay of profitability, lower required skill levels, and an expanding commission-based economy; recent examples have underscored the high financial attrition rates on the victim organizations. Since such attacks might prove damaging to specific companies involved in producing metals and minerals for the microsystems supply chain, these risks must be considered within the UAMY-CRIS perspective. Any cyber-warfare-related disruptions of mineral mining and processing are likely to have broad implications across all sectors of the economy. Cyber attacks may also be evolving toward a future capability for cyber-physical assaults that are yet to be realized at the large systems or geopolitically damaging scale.

The threat landscape and defense surfaces of the mineral extraction and metal-processing domains encompass those systems and processes directly connected to the critical infrastructure of communications and the Internet. With increasing intensity, the criminals executing these attacks have also moved into a commission-based model in which systems controlling specific aspects of mineral or metal production are rented out for fraudulent and profit-gaining operations. One of the main factors contributing to the success of these attacks has been the general lack of awareness of the cyber-attack threat and the low technology confidence within the management and operational communities. These arresting factors have been exacerbated during the recent global pandemic, which has limited the opportunities for training and familiarization with the hacker community and their tools. Cyber-attack mitigation strategies should be applied using the principles and practices associated with the concept of “defense-in-depth.” Effective risk management of critical cyber-attack threats associated with the specific operational environment can be achieved through a combination of disaster recovery plans and cyber-incident response plans.

3.1. Threat landscape for mineral extraction and processing

Risks and vulnerabilities associated with cyberattacks in the mining and refining cyber-physical-environmental domain are reviewed, with a focus on the cybersecurity aspects of cyber-physical systems.

The reliance of the global economy on non-renewable resources has placed the mining and refining, in particular, in attractive targets by the state-sponsored actors and hacktivists. Remote mining activities are both egressible to by hackers and physical intruders since these activities are frequently insufficiently monitored because they are lengthy and difficult to inspect [1,11-12]. The primary target in attacks by states on their opponent is often mines and mineshafts because they are dependent on natural resources whose economic foundation and military warfare rely on them. Mining and refining processes tend to use autonomous vehicles addressing the risks of manned transports of the harmful materials to and off the remote locations but now the atomic agents become susceptible to the threats of multi-layer assaults on their cyber-physical space. Attackers of mining and refining will set their eyes upon the data, physical and digital, capable of compromising the existing operations and production in the future. The attack surfaces will take into account risks related to the depletion of resources and the corporate social responsibility to the country, community, and company [13-15]. After the risk evaluation, the defensive-in-depth strategies are recommended to areas which are pivotal in the operations, including supervisory control and data acquisition systems, industrial control systems, environmental sensor networks and autonomous systems. Business continuity planning externalizes the residual risk and develops the response action plans..

3.2. Attack surfaces and defense-in-depth strategies

The wide scope of attack, which is a mining and refining, encompasses corporate IT, OT, and the additional supply-chain links to key operations like monitoring, control, and data analysis based on provably dependable hardware, software, and information. Since there will be deliberate and accidental incidents that interfere with the mining and refining practices, a comprehensive defense-in-depth approach will be required [16,17]. The owners of assets are in need of a sound cyber-physical security practice and vast enterprise risk management (ERM) in order to avoid disruption, loss of sensitive information, loss of intellectual property, trade secrets, or capital. All firms, regardless of size or industry that mine or are associated with the mining and refining of critical and valuable metals should have an elaborate incident-response capacity.

Critical mineral resources are those that are strategic and vital towards the smooth running of the economy of a given country. The origins of economic security include diversity of supply, health and stability of the economies that were considered to be desirable sources as well as resilient supply chains which are defined by its trustworthiness and cyber and physical-security postures. Redditch Strategic Services' classification of operational areas for critical-focused mining presents a high-level input into identification of attack surfaces associated with these functions.

3.3. Operational risk management and incident response

Operational technology cybersecurity for mining and mineral processing factories is not addressed in the U.S. National Cybersecurity Strategy. The threat of unpredictable cyber-attacks on automation control systems for these facilities is an especially serious concern because any disruption to mineral operations can initiate immediate and systemic shortages throughout the global supply chain for critical minerals. Yet no widely supported and proven collective defense methodologies exist to help operational technology cybersecurity defense teams detect, respond to, and recover from cyber-attacks against automation control systems in the mineral mining and processing sector. Mining and mineral-refining facilities can be considered components of the integrated asset management domain.

In March 2023, the U.S. National Cybersecurity Strategy designated $20 billion of resources, telecommunications, and electric infrastructure to be covered by supply chain security. The cyber risk management of supply chains may be characterized by particular issues because they may include several organizations and jurisdictions. Therefore, collective security measures are frequently perceived as a need and hopefully can be mutually implemented following such systems. The individual parties take their preparatory and proactive defensive capacities and rely upon the preparedness, vigilance and goodwill of interdependent partners in offering detection, alerting and mitigation of an incident and also supporting them during a crisis [12,18-20]. Nonetheless, the provision of these capabilities is particularly difficult in the minerals industry since the critical-mineral markets are volatile by nature and therefore, not able to sustain an investment on detecting and responsive capabilities on the part of any single company/ jurisdiction..

4. Semiconductor Supply Chain Protection through AI

Artificial Intelligence (AI) provides opportunities for protecting semiconductor supply chains against the most pressing concerns, especially lifecycle integrity, counterfeit prevention, and resilience It can facilitate multilayer traceability of electronic chips providing a solid foundation on vulnerability evaluation and mitigation, risk scoring, assurance segmentation, and authenticity determination of chips

with regard to source, integrity, and ownership. The warning and decision support in the chip supply, fabrication and packaging eco-system could also be enabled through threat-hunting AI.

Supply Chain as Traceability, Tamper Detection, and Counterfeit Risk Global Security requires robust traceability during the entire chip lifecycle, including design, chip supply, and integration into electronic equipment.. Traditionally, chip origin and integrity information has been contained within the IHS Markit Supply Chain Platform deliverable — for example, an integrated circuit identification and handling requirements — and supported by chips loaded with digital signatures [21-23]. Leveraging new sources and approaches can provide deeper lifecycle visibility and assist in countering increased chip counterfeiting and tampering concerns.Using natural language processing, an evidential knowledge graph of the chip supply chain — such as sources, assembly locations, supply, and integration partners — can be built and continuously maintained. A second knowledge graph can be filled and updated on risk intelligence information regarding counterfeiting and tampering, including technical description, warning messages by reputable organizations, as well as ongoing reports.

4.1. Chip lifecycle traceability and supplier integrity

Artificial intelligence can also strengthen security within the semiconductor supply chains due to provision of reliable suppliers, materials and components. The demonstration of chip-design elements, as well as the trace of the lifecycle of chips, is provided with the help of smart contracts, founded on distributed ledgers. AI is used to identify undoctored functional test data because any modification may suggest that a defective chip has been replaced in a product with a new one. The systems of chip-design, fabrication, packaging, and testing need to prevent the active application of malicious backdoors and side channels, and the threat of risk-management programs should include the expectation and countermeasures of them [24,25]. Such enabling actions as the observation of the creation of superior software to develop chip design and manufacture, finding suppliers of low-quality chip-test and packaging services, and the use of generative and predictive AI in the packaging process are in support.

The semiconductor complex involves a vast structure of suppliers as well as clients in the chip-designs services to assembly, packaging, testing and set-up into electronic structures. A vast majority of silicon wafers used in making logic ICs have Taiwanese and a significant amount of South Korean origin. The design of chips is done in a diverse number of countries, though most are made in East Asia, and a small number of the design and packaging have been achieved in the United States. The risk indicators of the supply chain differ according to the layer.

4.2. tamper detection, counterfeit risk, and verification

The chip components and the systems that they will be a part of integrity are critical. Tampering of chips in the production process is characterized by the insertion of extraneous circuitry to implanted chips or insertion of malicious components onto or into chips. The threat of producing chips in a manner that is purposely designed to serve the interests of an adversary a counterfeit production is valuable to protect against. Based on past compromises through so-called “supply chain attacks,” the assumption has raged that semiconductor production itself must not only be guarded heavily but should also be undertaken in a “trusted east” location. The introduction of AI-enabled generative adversarial networks (GANs) brings to the forefront the question of provable authenticity of chips and systems containing them in a radically new manner.

GANs are being developed to create photorealistic images of anything, including an arbitrary scene or the menu of an arbitrary restaurant in a specific photographic style. An embodiment could also create electromagnetic signals. Adversaries could use GANs to design electromagnetic signals or low-power hardware disguised as a highly trusted device to elicit close proximity access to an information system. Or an adversary could design a very detailed spoof neural network that tricked advanced deep fakes detection and adapted to an original content creator's style [26-28]. Therefore, GAN-influenced designs can quickly be fed into detection and generation–based detection models, and a prototype detection network could be trained with a limited dataset on high-performance computing hardware.

4.3. Resilience planning across fabrication and packaging

Planning resilience involves understanding single points of failure or factors that could cause widespread disruption to the supply chain. There are multiple potential reasons for disruption at design stage, both in terms of component location, health, and business resilience, dependencies (e.g., capacity within fabrication supply chain for tested bare dies for higher-end chips), the testing-lab business for tested dies of lower-end chips, and in the integration layer (known compatibility, etc.).

Resilience of assembly and packaging locations and suppliers (e.g., by using existing data on chips packaged in Taiwan and USA vs. elsewhere, and the choice of substrate suppliers, which chipmakers can also influence) must also be reviewed from a national-security point of view. These locations represent a single point of production for many security-relevant nodes, sometimes of very small size or using very novel packaging techniques [29-31]. Logically, the much higher volumes at OSATs also risk themselves much more to general incidents (be they natural, health, or CVE-related) and must plan accordingly. Simulation and generative AI technologies can point to strategies to enhance security and resilience of assembly-and-packaging-on-production throughput and overall supply-chain supply-and-demand balance.

5. AI for Energy Grid Resilience and Power Security

Electric power grids are two-sided cyber–physical systems that match energy demand with generation. Though the U.S. electrical grid still operates in near-real-time, the energy community anticipates much longer response times in the common electricity market under Time-Based Rate programs. New AI applications can assist utilities and elements of the power supply chain across multiple planning horizons and both operational faces of power grid operations. System models, weather forecasts and historical events can be trained on generative AI models like chatgpt to offer consultation to the analysts. Predictive AI can also be used in energy management, control, field and information technology systems to provide automated high-load forecast and proactive load-shifting commands to industrial and commercial customers..

Moreover, given the growing cyber- and supply-chain threats to the multi-stage semiconductor ecosystem underpinning communication technology, making the entire semiconductor supply chain more trustworthy is of increasingly high importance. AI can support the information necessary to enhance security against such threats throughout the semiconductor chip's lifetime, from research and development and design through field testing, use, and decommissioning. New probabilistic models enable chips to be traced over multiple stages. AI offers solutions to protect against forgery and counterfeiting and to enable the condition, performance, and functionality of chips manufactured and put into service by systems or networks to be confirmed at any time.

5.1. Demand forecasting, load shifting, and grid flexibility

The rising portion of demand of electricity can be easily adjusted to the supply availability. People and companies can be encouraged to prepare to change their carbon intensity or utilization of renewable by the use of AI to provide fine-grain estimates of future demand. Also available at the aggregate level, the electric power planners can use such information - in combination with the forecasts about the actual supply to determine when it is necessary to invoke relatively small scale load-relief or load-generation switching to stimulate the continued balance of the grid, AC frequency response during the emergency events and other physical issues.

Generative and predictive AI has also been put into use in improving operational planning and live monitor of the energy grid and associated networks. The transmission lines, substations, distribution systems, electric vehicles, and installations comprising battery-storages have all been put into consideration by a number of researchers exploring the use of AI-enabled decision support systems and procedures [3,32,33]. The operational research of cyber-physical infrastructure has been traditionally dominated by quantitative methods such as statistical estimation, optimization and simulation. Nevertheless, the advantages of generative and predictive AI such as fast generative design, simulationally replicatable stitches are also currently being harnessed to enhance operational decision-support of progressively cyber-physical energy grids.

5.2. Generative and predictive AI in grid operations

Managing distributed energy resources can be optimized via generative AI to promote management activities not present in the historical data through the production of new time-series data. As one example, a generative deep learning system can rapidly generate realistic time-series data to designate how the rooftop solar generation and the electric vehicle charging can be coordinated flexibly to lower retail power expenses. On the same note, offline generative AI may be used to inject exogenous variables into the historical distributions of the local electricity loads to train demand forecasting models that can inject noise in times of peak demand. The existing demand response initiatives can be enhanced greatly based on bigger datasets that represent a broader scope of situations.

The predictive AI models are usually data-intensive yet may be utilized in grid operations as well as to augment tuning in less data-abundant regimes. In one case, a Scottish utility has taken advantage of the shortage on operational data in 201819 and developed an electricity load forecasting system based on 123 attributes using multifactor analysis of numerous exogenous variables. This kind of model assists decision makers to make it through unserved energy that may arise even during uncharacteristic events..

5.3. Cyber-physical security and incident readiness

To curb the cyber-physical security risks associated with energy grids and provide a prompt and efficient response following the eventuality, AI-intensive functionality, in particular, demand forecasting, predictive analytics, and failure detection, should be utilized to promote the safety of cybersecurity and operational resilience. The large power systems (LPS) feature a large number of aspects that fall under the domain of cybersecurity, which means that optimizing them poses extra scaling difficulties, as it is highly coupled regarding the aspects of operations and energy, and offers radical new solutions.

The current source of data, e.g., AI-based surveillance of the social media intelligence (SMI), can be used to form demand predictions, and the ability of the AI to detect time and multimodal patterns can make the use of such methods to employ the use of varied streams of data. Load-shifting processes to increase the

balance between demand and supply can exploit the flexibility on the short timescales, or blurred price [4,34-36]. In the case of greater time scales, and in contingency planning, penalized optimization-based methods can be used to plan and mobilize contingency mechanisms like DR and storage deployment. Predictive capabilities at AI levels of subcomponent diagnostics, SCADA supervisory control and data acquisition, and end-to-end failure are developed to detect events in advance and to model propagation both of which can cause timely preparative actions on multi-dimensional levels.

An alarming number of cyber-physical incidents have occurred on LPS, yet AI technology has thus far had little impact on either prevention or preparation, even though these processes might be easily and beneficially supported by the technology. Operational risk management, liability considerations, and other related domains are therefore good candidates for exploration in terms of AI-enhanced support. AI-based signal processing can be employed to detect incursions in the physical space of LPS. Such capabilities can usefully be enriched through development of robust cyber-physical incident response plans, either as part of an overall scheme for incident preparedness in a prescribed area or with respect to any particular dimension.

6. Integrative Frameworks for AI-Enabled Supply Chain Security

Americas critical supply chain AI-enabled security is the solution-based approach to developing the synthesis of significant themes and includes such elements as governance and standards, technical frameworks of data interoperability and evidence-based measures of the performance of such initiatives by various industries and diverse stakeholders.

The existence of governance structures that put in place an adequately minimal adequate governance of any AI research initiative is a well-known concept. However, such a strategy is inadequately broad for the entirety of the AI ecosystem, short-circuits longer-range thinking on the crucial aspects, and yields governance instruments inappropriate for many AI tools. Indeed, as an example, an open-ended threat factor prospective study, which is akin to Canadian investments in GPT, shows that any intervention to avert such concerns means that public-private programs are far wider than the current malware detection and disinformation offensive campaigns. Governance broad enough to begin addressing these multi-faceted and multi-stakeholder strategies would benefit from identifying the main supply chain sectors, standardized components of the AI-enabled solutions, and strategic areas of attention.

Methods supporting the effectiveness of these supply chain security efforts are also nascent [37-40]. A key requirement is the establishment of data frameworks that combine the necessities for high-performance AI with the hidden desire for supervision, control, and unrestricted ownership. This leads to a new formulation of data interoperability: Interoperability that works for data and AI, but is designed for humans. It is non-functional in the AI sense because privacy and data sovereignty concerns are no longer shoehorned into technical solutions that sacrifice performance.

6.1. Governance, standards, and interoperability

Governance models facilitate the architecture of AI-assisted security of supply-chain within the industries. It is the establishment of trust-enabling standards as well as horizontal interoperability that facilitates the alignment of a wide range of stakeholders to the proposed strategy which results in successful implementation of vertical implementability in the end. The paradigms of trade, finance, and cybersecurity are well-known paradigms of governance-by-governance, which can be taken as an

example. They now become necessary in the sphere of AI-assisted security of supply chains of critical importance, in the ways of empowering vertical business ecosystems and horizontal partnerships.

To have security of the supply-chain using AI, multiple sectors with federation across businesses will be needed to collaborate. It relies on the capacity of a governance model, facilitating collaboration between industries, beginning with minerals, semiconductors, and energy, and collaboration of the public body, academia, and the civil society. The benefit of coordination is that there is minimization of risks, cost and effort by all parties involved [4,41,42]. It can be assisted by AI with the augmentation or even automation of the numerous governance processes, provided that people trust it and a set of ethical work is delivered.. The next step combines a how-to approach with a set of appropriate criteria. Avoiding hype and operationalizing a well-articulated agenda constitutes its foundation.

6.2. Data interoperability, privacy, and sovereignty

Enhancing organization and jurisdictions through sharing information on scale would support the securing of key supply chain segments and watershed sectors. This sharing is made possible by AI technology but the same algorithms and architectures also run significant risk should they be implemented without considering the issue of privacy, sovereignty and trust. The collective risk is de facto due to the data crowding and the widespread use of large foundation models in sensitive fields, where the information producers do not know that the advantages of sharing the information will not be received disproportionally among other stakeholders. The synthesis and analysis of deep fakes using statistical techniques offer novel problems of having to verify the provenance of the data and setting up trade-offs between quality, quantity, and data privacy. Technology ecologies integrating both high encryption with decentralized learning, e.g. federated learning, differential privacy, and homomorphic encryption reduce the risks of organization and jurisdiction but are resource and data intensive. The problem of these conflicting requirements needs to be solved transparently and overtly to supply chain security and, in fact, the adoption of AI in general.

The instrumentation used to align foundation model is still at an infancy stage and the practical implementation is mainly trial and error and close guidance of a human being. The effectiveness of machine-made advice on the professionalism of specialists is immeasurably complicated. Even whether predictive or generative AI will be better or worse to the society in the long term is not that obvious. The human attention further complicates the issue since it is never easy to foresee whether and on what circumstances generative AI will be encouraged to produce more valuable, meaningful, and relevant content than it impregnates on human sources [43-45]. The long-term knowledge ecosystem implications of foundation model articulation or misinformation are not only difficult to predict and assess, but also live in a scientific and political limbo, lacking the resources or organizational structure for proactive resolution.

6.3. Evaluation metrics and evidence-based policy implications

Concrete evaluation metrics are critical for promoting successful AI-enabled supply chain security and resilience initiatives, and supporting broader investments in trust, security, and privacy by addressing the public-good character of data security. Many of the current AI resilience initiatives in these sectors are nascent, with unproven return-on-investment, and the community must be cautious about applying generic commercial-off-the-shelf productivity metrics that do not take into account the specific public-good nature of many of the investments, looking instead for composite evidence that builds confidence in

the effectiveness of the AI interventions. Caution is also warranted in selecting the AI capabilities and technologies that appear to be paying off. The productivity increase due to human capital and robots has been hugely exaggerated by analysts, and probably the two factors also represent only a delay effect on falling aggregate multifactor productivity growth. Supply chain security and resiliency investments powered by AI are across several nontelling sectors, where policymakers at earlier stages have been more hesitant to act, and the system of data production and platform status are not sufficiently developed but rather being developed and deployed, not utilized. The standard delivery for the market is still stability and security.

Evidence-based policy elaboration can also benefit from a closer qualitative, sector-by-sector examination of the underlying characteristics of different classes of AI applications, shaped by concepts like Moore's Law, likely exhaustible productivity sources, and their relation to specific public-good characteristics of the AI support infrastructure [9,46-48]. Research has shown that breakthrough-altering technologies, from steel to railroads, demanded profound reframing of the dominant socioeconomic entities, organizations, and institutions governing and supporting the economy, and have not come from simply letting the market operate multipliers in sectors and economies where such factors have hitherto been hidden up pebble bathtubs.

7. Case Studies and Implementation Roadmaps

Protecting American semiconductor supply chains against high impact cyberattacks has been adopted at national level. Even after investments in domestic and overseas fabs that are subsidized, the security of the upstream fabrication and packaging of the business is posed as a security threat. It presents AI methods based on the protection of the integrity of products with the threat of mitigating the success of attacks and ensuring resilience and readiness to incidents using advanced persistent threats as a frame.

The semiconductor technology is so common in the contemporary society and the stability and integrity of the chip hardware is a critical factor. The integrity of the supply chain is an issue that runs down to design and fabrication of products to the packaging of the final products. The use of AI is brought out as the facilitator of clear risk management, particularly in the cases of latent malice like Trojans, backdoors, counterfeiting, and attack in the packaging process. The attack area is big, and lifecycle lifecycle defense is required in order to minimize risk.

Openness and worker honesty by means of generative AI-based joint distribution modeling at scale identifies the sheer levels of dependency in the ecosystem and the opportunity to evaluate the likelihood and effect of particular suppliers to be succumbed along certain parts. Predictive risk indication and visualization which is AI-assisted analysis of shielding efficacy and foreshadowing predict susceptibility to tampering damage. Multi-relationship analysis, integration of the environmental conditions using AI and share-based monitoring at those suppliers of major products assist in general securing the supply chain through the reliability aspects on components.

7.1. Minerals and mining sector deployment

Having to endure the geopolitical tensions, remote COVID-19 and naked weather extremities caused by the climate change, the United States is starting to see the necessity to take immediate and firm steps to reduce the national video threat of critical supply chains in America. Almost every contemporary mass-produced product and many services depend on critical minerals which is a list of quantifiable commodities rare earth elements (REEs), tin, lithium, cobalt and copper, the supply chains of which are

subject to disruption. Other industrial sectors that are easily accessible and vulnerable to manipulation are mineral mining and the treatment process; informing adversarial cyber activity objectives towards the lifelines of the U.S.A.

The two broad types of activity targeting critical minerals include further development of the underlying tracking schemes and data provenance operations informing the risk assessment, and use of risk monitoring and management machinery to offer a radiographic and evaluate potential vulnerability of the supply chain in the critical minerals production and refining community. The same issues regarding the ethical, legal, and environmental (ELE) aspects peculiar to mining and mineral processing are reflected in both applications, and their integration implies both the need to address the concept of supply chain security on an agency-wide level of the whole manufacturing environment, and the benefits of an integrated and sector-agnostic approach.

7.2. Semiconductor ecosystem protection

Semiconductor industry ecosystem is critical to the national security and economical wellbeing and it experiences increasing supply chain risks due to geopolitical disturbances. The target in making the semiconductor supply chain safer is to rely on AI and machine learning functions to have improved traceability, integrity, security, utility and responsiveness of the overall semiconductor chip life cycle, encompassing design, wafer manufacturing, assembly/ testing, integration, and deployment. New advances in generative AI form the basis of ways of enhancing security against tampering, counterfeiting, back-doors and other forms of integrity assurance in both chips and systems. Techniques of defense- in-depth and deploying strategies to restore functional integrity in case of an incident on the supply chain parties either as a result of sabotage, mishap, attack, or natural disaster are also being formulated..

Existing C5I-enabled infrastructure has delineated the attack surface for the U.S. semiconductor ecosystem and provides for the proactive integration of Commercial Facilities cyber-physical security into C5I-wide and HSPD-12-related defense planning. Existing AI capabilities can also play important roles in helping regional and intra-ecosystem market actors prepare and reduce resilience-relevant dependencies on single suppliers [33,49]. This focus has been informed and guided by the subject matter expert workshops at MITRE, such as those in the MIET and MSI, the subject of which are semiconductor supply chain security and integrated cyber-physical security.

7.3. Energy grid resilience initiatives

The issue of power generation and transmission should be regarded as a national security requirement and current strides are aimed at creating a supply chain of electric vehicles and batteries which can be considered as a close-term strategy. Setting up of electricity supply should however be well planned and secured also. Demand planning is needed since the load should be tightly matched to the generation at all times, causing expensive cutbacks or disruption in case of over load forecast and over the required generation capacity. Planned demand-shifting activities can also help to maintain grid resiliency at times of extreme weather conditions as well as unforeseen infrastructure failures. Besides these load-shifting functions, broad-spectrum flexible power systems are also attracting predictive systems based on model-based AI applications or, data-driven AI applications. The climate acts positively push or instigate more possibilities of glitches and assault on these essential infrastructures. The inclusion of a cyber-physical level of security and the preparation of incidents is consequent.

Demand and supply of electricity have to be appropriately balanced at any time and security repercussions of deviation should be reduced. The grid resilience can be supported by AI solutions that enhance the demand forecasting and apply smart demand supported by adequate residential energy flexibility. Various grid-operations tasks, including incident avoidance and diagnosis, have a major potential with the help of generative and predictive AI. A cyber-physical defense layer can also be boosted by generative approaches that facilitate the increase of attack- and incident-preparedness levels. The interfaces facilitate interests of UAMY by lessening operational effects of safety and security events and considering the propagative damage past the initial system.

8. Conclusion

The focus on the governance, policy, standards and investment levels of the AI-enabled security solutions is necessary as it is intersectoral and inter supply chain and must be paid close attention to in order to ensure that the results are the best they can be and bring only the most desired changes.. Assurance-based third-party governance and interoperability standards are necessary for creating qualitative and quantitative AI-AF solutions. Data inference and AI provenance frameworks are needed to help ensure consumers' rights without compromising supply chain integrity. Clear evaluation measures can achieve the efficiency of investments in AI-enabled supply chain security through a trade-off mechanism conducted across different policies. The case studies provided show a plan of implementing AI-AF solutions in the minerals and mining industry, semiconductor ecosystem, and the energy grid.

The best that can be done is to provide domestic supply-chain resilience, knowing that it is not an easy task due to deeply rooted social, industrial and economic factors. In the short run, there is a viable operation strategy by minimizing risks in sources and ecosystems of susceptibility in the supply-chain.. Towards this goal, the AI-AF approach aims to reduce risks of foreign dependence, catastrophic threat exposure, and demand surge; and to support incident investigation and mitigation. Enabling technology is viewed as a laptop computer with enhanced Internet connectivity and zero-dollar software-from-the-cloud support for citizens-and compassionate governance.

References

[1] Abdelwahab HR, Rauf A, Chen D. Business students' perceptions of Dutch higher educational institutions in preparing them for artificial intelligence work environments. Industry and Higher Education. 2023 Feb;37(1):22-34.

[2] Alam A, Mohanty A. Business models, business strategies, and innovations in EdTech companies: integration of learning analytics and artificial intelligence in higher education. In2022 IEEE 6th Conference on Information and Communication Technology (CICT) 2022 Nov 18 (pp. 1-6). IEEE.

[3] Mohapatra PS. Artificial intelligence and machine learning for test engineers: concepts in software quality assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:17.

[4] Allal-Chérif O, Simón-Moya V, Ballester AC. Intelligent purchasing: How artificial intelligence can redefine the purchasing function. Journal of Business Research. 2021 Jan 1;124:69-76.

[5] Amirkolaii KN, Baboli A, Shahzad MK, Tonadre R. Demand forecasting for irregular demands in business aircraft spare parts supply chains by using artificial intelligence (AI). IFAC-PapersOnLine. 2017 Jul 1;50(1):15221-6.

[6] Badghish S, Soomro YA. Artificial intelligence adoption by SMEs to achieve sustainable business performance: application of technology–organization–environment framework. Sustainability. 2024 Feb 24;16(5):1864.

[7] Bickley SJ, Macintyre A, Torgler B. Artificial intelligence and big data in sustainable entrepreneurship. Journal of Economic Surveys. 2025 Feb;39(1):103-45.
[8] Cavazza A, Dal Mas F, Paoloni P, Manzo M. Artificial intelligence and new business models in agriculture: a structured literature review and future research agenda. British Food Journal. 2023 Jul 12;125(13):436-61.
[9] Panda S. Scalable Artificial Intelligence Systems: Cloud-Native, Edge-AI, MLOps, and Governance for Real-World Deployment. Deep Science Publishing; 2025 Jul 28.
[10] Chen R, Zhang T. Artificial intelligence applications implication for ESG performance: can digital transformation of enterprises promote sustainable development?. Chinese Management Studies. 2025 May 13;19(3):676-701.
[11] Demaidi MN. Artificial intelligence national strategy in a developing country. Ai & Society. 2025 Feb;40(2):423-35.
[12] Drydakis N. Artificial Intelligence and reduced SMEs' business risks. A dynamic capabilities analysis during the COVID-19 pandemic. Information Systems Frontiers. 2022 Aug;24(4):1223-47.
[13] Fallahi S, Mellquist AC, Mogren O, Listo Zec E, Algurén P, Hallquist L. Financing solutions for circular business models: Exploring the role of business ecosystems and artificial intelligence. Business Strategy and the Environment. 2023 Sep;32(6):3233-48.
[14] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.
[15] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[16] Fonseka K, Jaharadak AA, Raman M. Impact of E-commerce adoption on business performance of SMEs in Sri Lanka; moderating role of artificial intelligence. International Journal of Social Economics. 2022 May 16;49(10):1518-31.
[17] Gursoy D, Cai R. Artificial intelligence: an overview of research trends and future directions. International Journal of Contemporary Hospitality Management. 2025 Jan 2;37(1):1-7.
[18] Hu KH, Chen FH, Hsu MF, Tzeng GH. Governance of artificial intelligence applications in a business audit via a fusion fuzzy multiple rule-based decision-making model. Financial Innovation. 2023 Aug 14;9(1):117.
[19] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.
[20] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.
[21] Jorzik P, Antonio JL, Kanbach DK, Kallmuenzer A, Kraus S. Sowing the seeds for sustainability: A business model innovation perspective on artificial intelligence in green technology startups. Technological forecasting and social change. 2024 Nov 1;208:123653.
[22] Jorzik P, Yigit A, Kanbach DK, Kraus S, Dabić M. Artificial intelligence-enabled business model innovation: Competencies and roles of top management. IEEE transactions on engineering management. 2023 May 24;71:7044-56.
[23] Kalogiannidis S, Kalfas D, Papaevangelou O, Giannarakis G, Chatzitheodoridis F. The role of artificial intelligence technology in predictive risk assessment for business continuity: A case study of Greece. Risks. 2024 Jan 23;12(2):19.
[24] Krishnan C, Gupta A, Gupta A, Singh G. Impact of artificial intelligence-based chatbots on customer engagement and business growth. InDeep learning for social media data analytics 2022 Sep 19 (pp. 195-210). Cham: Springer International Publishing.
[25] Lada S, Chekima B, Karim MR, Fabeil NF, Ayub MS, Amirul SM, Ansar R, Bouteraa M, Fook LM, Zaki HO. Determining factors related to artificial intelligence (AI) adoption among Malaysia's small and medium-sized businesses. Journal of Open Innovation: Technology, Market, and Complexity. 2023 Dec 1;9(4):100144.

[26] Met İ, Kabukçu D, Uzunoğulları G, Soyalp Ü, Dakdevir T. Transformation of business model in finance sector with artificial intelligence and robotic process automation. InDigital business strategies in blockchain ecosystems: Transformational design and future of global business 2019 Nov 10 (pp. 3-29). Cham: Springer International Publishing.
[27] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.
[28] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[29] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.
[30] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.
[31] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..
[32] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[33] Qi B, Shen Y, Xu T. An artificial-intelligence-enabled sustainable supply chain model for B2C E-commerce business in the international trade. Technological forecasting and social change. 2023 Jun 1;191:122491.
[34] Sahoo S, Kumar S, Donthu N, Singh AK. Artificial intelligence capabilities, open innovation, and business performance–Empirical insights from multinational B2B companies. Industrial marketing management. 2024 Feb 1;117:28-41.
[35] Saleem I, Al-Breiki NS, Asad M. The nexus of artificial intelligence, frugal innovation and business model innovation to nurture internationalization: A survey of SME's readiness. Journal of Open Innovation: Technology, Market, and Complexity. 2024 Sep 1;10(3):100326.
[36] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[37] Shaik AS, Alshibani SM, Jain G, Gupta B, Mehrotra A. Artificial intelligence (AI)-driven strategic business model innovations in small-and medium-sized enterprises. Insights on technological and strategic enablers for carbon neutral businesses. Business Strategy and the Environment. 2024 May;33(4):2731-51.
[38] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.
[39] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.
[40] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.
[41] Shivadekar S, Mangalagiri J, Nguyen P, Chapman D, Halem M, Gite R. An intelligent parallel distributed streaming framework for near real-time science sensors and high-resolution medical images. In50th International conference on parallel processing workshop 2021 Aug 9 (pp. 1-9).
[42] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.
[43] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.

[44] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.
[45] Tavera Romero CA, Ortiz JH, Khalaf OI, Ríos Prado A. Business intelligence: business evolution after industry 4.0. Sustainability. 2021 Sep 7;13(18):10026.
[46] Upadhyay N, Upadhyay S, Al-Debei MM, Baabdullah AM, Dwivedi YK. The influence of digital entrepreneurship and entrepreneurial orientation on intention of family businesses to adopt artificial intelligence: examining the mediating role of business innovativeness. International Journal of Entrepreneurial Behavior & Research. 2023 Jan 17;29(1):80-115.
[47] Wach K, Duong CD, Ejdys J, Kazlauskaitė R, Korzynski P, Mazurek G, Paliszkiewicz J, Ziemba E. The dark side of generative artificial intelligence: A critical analysis of controversies and risks of ChatGPT. Entrepreneurial Business and Economics Review. 2023 Jun 30;11(2):7-30.
[48] Zhou X, Li G, Wang Q, Li Y, Zhou D. Artificial intelligence, corporate information governance and ESG performance: Quasi-experimental evidence from China. International Review of Financial Analysis. 2025 Jun 1;102:104087.
[49] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.

Deep Science Publishing, 2025
https://doi.org/10.70593/978-93-7185-637-9

DeepScience
Open Access Books

Chapter 10: The Road Ahead: Building an AI-Secure America

1. Introduction: Vision for an AI-Secure United States

An artificial intelligence secure United States uses the strength of artificial intelligence to establish a stable, wealthy, and just country and the rest of the world. AI is an up-and-coming support of the U.S. economy alongside cybersecurity, climate resilience and medical care.One can establish defendable growth and gain employment in these fields. The society is relying on the benefits of AI, one of the pillars of technology that have driven autonomous cars, smart chatbots, faster drug discovery, and so on, to counter the harms of AI, such as algorithmic bias, threat to the safety of society, unemployment, antisocial technologies and other shadow aspects. Protecting students, employees, purchasers, and the entire society, in general, puts the United States on track to an AI-driven world.

The vision draws upon self-supporting investments in research, regulation, and innovation, economic cooperation, education (inter-sector, and with allies) along the economy-executive education-lifelong learning pipeline. Weaknesses jeopardize the achievement of the chance of the moment. AI builds the future of labour and reduces the shortage of skills through reskilling, retraining, and upskilling at the same time. In its right hands, the stigmatization can be wandered to an extent of bringing back lost personnel to the labor force. It is important to tackle these risk and uncertainty sources in order to establish a stable growth course towards the United States.

2. Workforce development and AI education

Policymakers should focus on reinforcing the education pipelines and encourage lifelong learning. Although the demand to have AI-skilled labor is vast, the situation between K-12 education and university graduation is still unsatisfactory. In addition, existing employees are equally to be impacted on their career by AI yet most disciplines cannot offer them meaningful and novel scalable retraining programs [1,2]. The increasing disparities in the base knowledge of younger generations, particularly in mathematics and computer science, have stranded many applicants wrongly equipped to take jobs in AI, and will continue to reduce the AI talent pool in the US even in the years to come. To make the United States be well-positioned to compete successfully in an AI-enabled economy, the country must not just further invest in technical infrastructure of AI, but also offer reskilling to employees in AI-related professions. The combination of these efforts will aid in hiring and training the most qualified AI human resources.

The policymakers will be able to contribute to strengthening education pipelines and reskilling. That would necessitate closing gaps in the foundations of K 12 education, setting standards of capabilities in related jobs in the field of AI, growing and cultivating covert recognition in collaboration between the government and business in expertise cultivation and making lifelong learning a national priority. The

investment in this area will enable the labor market to facilitate innovation in the coming days. On the K-12 level, the knowledge of AI roads models and prompt-response pairs within the scenarios of normal scale and tolerance. A high-school degree will require cursory knowledge of models of language and vision, a ability to create simple prompt sets, and practice in processing the results of a model. On the adult-learning level, courses focused on managers in AI-related sectors should allow identifying the abilities to produce texts, images, or audio, and training in technical-advisor should include how to implement the abilities to the team, process, and toolset and include the basic ethics and bias-reduction knowledge..

2.1. Education pipelines and lifelong learning

To have AI-safe America, gradual improvements to talent pool of the country, both in the robust pipelines to AI and re-skilling prospects of the adult population, are required. It should be paid attention at every level of education. Though the main problem of stakeholders reacting to such discourses is higher education, an AI-driven economy requires the mechanisms of developing talent starting at K-12 and continuing at graduate and further education. Simply connected with it is the issue of lifelong learning, which has gathered the growing interest of bipartisan attention [3-5]. The objectives of such efforts include to offer upskilling and re-skilling opportunities to the adults, to make sure that the initial programs will instill the initial skills without which the lifelong education will not be accomplished successfully.

The solution of these objectives is a complicated task. The problem of skills shortages is not new, neither is the regular demand of the programs facilitating the growth of training, working, and development of both the public and private training and education sectors. The difference now, besides the unparalleled rate of technological development, is the multifactor quality of the changes that have happened in the needs of that labor market. Numerous AI features are technical certifications in fact which indicate skillfulness in a small area. Unless the right coordination is done, one institution can end up spending time and efforts on a credential that the potential employers do not recognize. Existence of responses should be consequently far greater in scope and partnership than before programs. The list of some of the examples mentioned is a proof of concept; by no means are they a comprehensive solution.

2.2. Skill benchmarks and credentialing for an AI-driven economy

The competency standards and certifications in an AI economy should be made in such a way that the skill levels can enable workers to adapt to the fast changing needs of the market. This entails the articulation of the educational outcomes, identification of skills that cut across the changing job roles and the alignment of skills supply with the demand of the employer [6,7]. The results must not only be stated in terms of desired skills, but the means by which development and achievement might be confirmed within a decentralized, flexible setting.

The skill benchmarks are identified with particular abilities and knowledge required in the entry-level and advanced jobs within an AI economy, based on the models and talent-information systems that are outlined in two sections above. The skills and knowledge required in a particular AI-adjacent job role can be characterized by metrics at five levels of complexity and the level of analytical rigor required to be applied at each stage of the process would rise, Six-Stage Models of AI-Driven Job Roles; AI Domain Domains; The AI-Related Skills Landscape. To its credit, all of these sources can be pumped into a

supplydemand paradigm in the following way: Realizing the Vision of AI in the U.S. Economy; Dynamic Skill Demand; Supply and Demand: A Using an Enabling.

These benchmarks should be checked at every level by the credentialing systems. In the end, skills would be authenticated as per the employers, good educational providers, apprenticeships, and accepted professionals, in a decentralized and fragmented verification and certification environment. Such supply might be linked with real time demand through systems such as the National Skills to Job platform which would offer vital input of the whole skills-outcomes system by the private-sector.

2.3. Public-private collaboration in talent development

The training of AI talents implies wide-ranging public-corporate cooperation on the depths and levels of the learning ecosystem. These alliances need to share risks and rewards well and stimulate individual contributions by each of the partners [2,8-10]. The government units must lay out the requirements of the skill requirements, allow funding and bear and certify training to companies and institutions that are coming into the skills-development gap. There are issues related to speedy feedback of pipeline performance by industry partners, as well as the sharing of recruitment and utilization information which is useful in the further refinement of the program. Dual missions in teaching and research to the benefit of the workforce-training infrastructure should be built up by the public universities and colleges.

Government foundations must330 encourage National Science Foundation funding of further development of current educational and workforce-development initiatives to bring size. The areas that are priority should be amplified with successful initiatives found at or around maturity. Their implementation with pilots and later appraisals as shown by tuning is rapidly enhanced by their positive local effect. Foundations can also help through the relationship through sponsorship of the sectors of the economy that have an incentive to give them such as the big tech, defense contractors and semiconductor manufacturers and any other organization that relates to AI.

3. Policies for national AI governance

Another factor of achieving an AI-secure United States is that regulatory clarity should be established and a credible federated model of AI governance developed. The targeted model consists in finding universal jurisdictional circles that help identify the most appropriate government level to control the use of AI systems under a particular situation and coordinate functionally comparable norms at state and federal levels.

It is also important to develop an ethics and accountability framework that the stakeholders can rely on to comprehend, reduce, and remediate the algorithmic biases and the execution of procedures and tools that geared towards supporting security-by-design and supply-chain integrity within the AI systems. Moreover, ensuring that the set of policies of the US complies with or at least does not conflict with the policies of allied nations of America will allow establishing a minimum level of collaboration of countries in AI management over which their cooperation will multiply.

3.1. Federated governance models and regulatory clarity

Regulatory clarity guidelines can help in realizing federated governance model particularly following the expansion and changes in jurisdiction. With the increase in the prevalence and capability of AI, it becomes an important spatiotemporal factor, an extra dimension of physical space, as some have described cyberspace, with corresponding objects and phenomena becoming salient and meaningful to

human stakeholders and intended action. As physical trade routes, resources, diseases, and weather are influenced, human and physical resource, process, dynamics, and trade or safety process boundaries breach jurisdictions. When trying to bring federal involvement and regulation in early efforts were made, they tried to harness these capabilities responsibly, without incurring a pareto-optimal opportunity cost which could cause the use of these new technologies. Having distinct demarcation of the territory of ethical use and consideration, therefore, entails coordination and collaboration among different jurisdictions which approach work in a federated way.

The harmonization of requirements of risk assessment, equity and bias reduction, and interpretation of results, bring the jurisdictions and organizational style towards the common objectives. Supply-chain maturity also requests services to operate in secureran environments throughout the life of the services and those dependent on them not only during the testing of the prototype but also during deployment. Both these constraints may demand a lot of organizations to auditing, particularly service-providing agencies, and should therefore be symbiotic to the federated governance.

3.2. Ethics, bias mitigation, and accountability frameworks

The healthy and trusted AI ecosystem is provided through the use of ethics, bias-mitigation and accountability frameworks. They establish limits of tolerable AI and make sure that the systems are safe and fair in their functioning [1,11-12]. Audit trails check on compliance and create accountability and failures can be solved promptly and parties benefiting can be made to be compensated on time.

The ethical boundaries of AI use ought to be installed with clear proposals specifying the method of holding the systems responsible to what they should consent, use and their decision-making. The developer, and the deployers of AI should be responsible in case of any bias, discrimination, or error of the system. Moral hazard in AI technologies that can perpetuate their creation, application, and use can be allowed by letting harm befall AI users or the systems themselves.

The government agencies must first take the initiative of engaging a multi-stakeholder group to put forward a series of ethical principles that can be used to implement AI in their respective jurisdictions and the applications that will be deployed by their agencies. The principles are supposed to focus on the nature of acceptable, responsible AI: that it requires clear governance mechanisms; that it not only needs enough variety in the data and individuals who will be involved in development, but also that it is expected that the outcome of the development will be constantly checked to understand whether it depends not on prejudices and can be true [13-15]. These principles must also point out departments and offices in charge of every ethical characteristic and quantitative measures to determine a success. These proposals then should be reviewed by an independent cross-agency committee and a coherent, unified set of values developed on all the federal government AI.

Multi-stakeholders groups should also collaborate with independent AI systems, such as judiciaries, regulatory bodies in the medical field, and police oversight councils, to create rules on how AI can be used in those areas. These groups also ought to be provided with relevant support either technical or financial to facilitate them perform effectively these mandates.

3.3. Security-by-design and supply chain integrity

In achieving the appropriate balance between other performance measures and in developing confidence in users and the wider society, it is important that security-from-design is ensured during conception,

development and deployment of advanced technologies. As AI capabilities advance and widen their foundation of application, their impact on the security of cyberspace, physical infrastructure, and cognitive processes is correspondingly deepened. The AI-aided actions of nation-state adversaries, terrorists, and cybercriminals are becoming ever more powerful and difficult to attribute, anticipating a future in which security may be for sale at scale—or given away in a competition-based economy. Shortcomings or malicious functionality introduced into a single system can be exploited at multiple additional targets due to data and model-sharing. Advancing AI-based types of attacks, attack plans, and to accelerate commodity access reduce the amount of personnel, finances, and skills needed to carry out advanced attacks [16,17]. The AI-based defense in its current form should therefore incorporate the new categories of mainly destructive and more novel attack into the conventional detective variables of primary and secondary adversary targets.

Risk management frameworks and supportive regulations providing guidance—without the inefficiencies of top-down prescriptions—are needed for critical infrastructure AI application systems and for the also-critical AI operation systems of untrusted suppliers, especially for consumer products. Detection using AI results in fresh methods of responding to incidents, predictive intelligence, and generation of risk-aware models; a blend transforms defensive models and tactics. The absence or inadequate quantitative measurements and practices are still seen as the obstacles to implementation in the key areas, and the direction to follow has not yet been determined. The same changes call for improved safety and verification-by-design practices in practical application areas: advice to an expanding audience about what should be changed in research funding processes to facilitate faster, deeper, safer development and deployment of AI, metrics for AI-enabled cyberattack detection and attribution, a more tractable model of stakeholder alignment for both AI-related protection and adversarial model sharing, and an alternative approach to insurance for highly dynamic systems.

3.4. International cooperation and standards

The cooperation and standards of different countries are crucial in ensuring that the potential of AI is used meritfully as security threats related to its rapid expansion are being addressed. The creation and use of AI are cross-jurisdictional and multidimensional processes, which are transacted on a global level. It is necessary to create an interoperable regulatory context including national and regional regulatory measures and policies, and even subsidiary national regulation, to carry out successful cross-border collaboration, and patchwork of regulations crafted by different nations and the creation of regulation frameworks on various levels may be a pitfall.. Standards and norms enable the emergence of best practices; once developed, they can help set governance, shape actions, and help determine desirable outcomes while promoting the trustworthy use of the technology.

The United States should invest in, promote, and drive international coalitions and partnerships to develop norms and standards associated with critical areas such as security, technical safeguards, trust, transparency, data sharing, robustness, malicious use, bias mitigation, performance measurement, and more. New alliances with the European Union, United Kingdom, Canada, Japan, Australia, New Zealand, NATO, G7 and other like-minded allies with similar values and goals are to be established with the aim of coming up with joint evaluation systems and best practices.. In addition, the United States should advocate establishing a shared global model for expression systems and models and encourage the multilateral, multipartite development of key security capabilities around those systems, beginning with the establishment of a global federation of cybersafety and AI safety labs [12,18-20].

4. Independent innovation ecosystems

The performance of a society can be correlated to the degree to which it relies on an independent and well-functioning innovation ecosystem. Autonomy in research transformation depends on the interdependencies with research ecosystems of other regions; whether these interdependencies are built on trust, openness, and collaboration; and whether they meet a set of guiding principles that ensure common interest and consideration of possible associated dangers. Such interdependencies exist not only among academic institutions, but also across the public and private sectors, the financial community, and small-, mid-sized-, and large-enterprise development.

The guiding principles for a healthy independent innovation ecosystem support the need to build a healthy internal research and funding ecosystem, safe the authorities' and institutions' capacities, and take a leadership role in regional funding mechanisms, enabling the focus on addressing the region's underlying needs and priorities, as well as on fostering the emergence of select flag projects [21-23]. A united approach to defining standards and processes that protect core intellectual property and ensure responsible sharing and replication is also essential for ensuring that private funds contribute to both the private and public interest.

4.1. Research autonomy and open science

In order to remain technologically ahead, the U.S. has to support autonomy in research, encourage responsible sharing and replicating. The Strategy acknowledges the value of American global scientific leadership and points out that alliances with reliable global allied partners can help in achieving these two objectives. According to representative and more senior technical staff, partnerships with other people outside of the U.S. tend to add value uniquely, such as providing them with unique data sets and knowledge elsewhere created or isolating them against certain domestic threats.

However, to retain the one-of-a-kind appealing companion, the U.S. has to ensure a friendly atmosphere to cross-border travels, cross-border exchange of goods, information and services, and the free flow of research and educational ideas among themselves. More elaborate application and review procedures which are risk adverse and open-ended as already cautioned by previous researchers would drive away valuable collaborations with non-allied countries [24,25]. The importance of science in the relationships between the U.S and China implies that it will be prudent to reduce the costs of doing business with reliable Chinese researchers through friction. More intense areas in research where the U.S. leadership can be successfully maintained, as well as the instances when alliances were particularly useful, should hold priority.

Openness and the idea of free-flowing cross-border research have been central to the U.S.-led global research enterprise. The resulting unusual level of share-then-check replication should therefore be sustained in AI and its allied fields through heightened awareness of good research practices, improved undergraduate education for PhD students, targeted funding for direct testing and development of urgently needed new replication protocols, and better funding for replication studies of Major Questions. The more extensive use of empirical testing data in proposed models is also useful. These life-critical propositions aim at making bias limiting, and making people test assumptions behind the assumptions, especially checking on the validity of the assumed relationships between inputs and outputs..

4.2. Federal and regional funding mechanisms

The vision statement of the United States government the one discussed on how to secure the future of the country with AI is going to need a long-term investment in the underlying research ecosystem. The funding systems that are organized at the federal and regional level should contribute sufficient resources at all phases of the research-to-market pipeline, but the number of funding priorities that can be named to correspond to significant problems of the AI Supply Chain are underfunded according to the recent legislation, especially at the early research stages [26-28]. These gaps can be addressed by competitive grants capitalizing on the quality of research universities in America along with sensitive-seed programs, consortium programs or prize-based programs targeting regional and cross-sector needs of AI Innovation, resilience and practical impact.

In the same way that the historic National Defense Education Act of 1958 did not only fund research, but also aimed to ensure the pipelines of STEM-trained talent and provide more funding and specialized resources to regional institutions and new interdisciplinary programs, modern-day investments in AI should create incentive to develop talent and AI Capability Development in the full panorama of universities, community colleges, trade schools, and Federal training agencies and give new impetus to implementation of innovative AI solutions within the public sector.

4.3. Entrepreneurial ecosystems and small- to mid-sized enterprises

U.S. government support for entrepreneurial ecosystems, and especially for non-monopoly small- to mid-sized enterprises, should seek to materially increase (i) the development of world-changing technology and its application in problem solving in all sectors of the U.S. economy, (ii) the creation of high-paying jobs, and (iii) systems that reduce stress while providing low-cost products and services to all citizens. This support should align capital, mentors, and regulatory relief. Much of the support should be regional in orientation, to better match physical social connections. By emphasizing technology creation and application in general rather than AI or even cyber security specifically, the solutions to other sectors' difficulties will become visible, thereby ensuring that the cross-sector connectivities remain live and functional. Firm size should not be a precondition for Federal funding or for experiments. Access to financially enabling capital markets is a primary deficiency in the US economy today; Regaining the ability of small and medium-sized firms to build sustainable companies while accessing IPOs will be very beneficial.

The over-centralisation of the market economy has triggered a consolidation of firms so unnerving that many policy analysts believe the innovation engine is failing. Without structural repair, public sector institutions will need to do the risk-enabled and subsidising functions that healthy competition normally triggers [29-31]. Pockets of innovation in various industries continue to flourish, but a picture of broad-based, sustainable, and structural innovation is absent. An AI-enthused excitement for ambitious solutions on a multitude of fronts should support those needing venture-level finance, particularly for revenue-concept companies pursuing USPTO and FDA approval. Dangerously, large global players in nearly every industries are proclaiming APRA a necessity; risking the possibility that the very creation and application of technology in these high offices will be lost.

4.4. Protecting intellectual property while fostering collaboration

Arguing about AI security often provides the answer that raises the question: how could intellectual property (IP)- which is so instrumental in building and maintaining the world-class innovation economy

of America- be safeguarded without suffocation of the augmented degree of collaboration that would be necessary in designing secure AI systems? The response should be subtle: a variety of institutional designs should be acceptable, self-censorship out of national security is not assailable activity, non-invasive (safe) benchmarking and audit to check the ethical and safety vigilance, and transparency in the process to warrant the centsuality of algorithms is necessary [3,32,33]. The operation of local innovation ecosystems based on AI may also be influenced by international interoperability whereby parallelity in demands on the collaboration of private sector is best being minimized to prevent the risk of under-investment in innovation and allowing selective and consistent over-provision of public good.

Three paths can be envisioned: the unanticipated qualities and capabilities of deep learning as national DARPA-like resources require cross-institutional sharing and collaboration to an unprecedented scale; the extreme complexity of knowledge-intensive AI systems requires an increasingly convoluted ecosystem of partners for each new generation, inevitably leading to concept and process leakage; and the sheer pace of progress—made paradoxically possible by the greater conservatism of larger players with access to market-leading equilibrium-driven datasets and the attendant fall of traditional AI capital markets—requires a more modular and decentralised organisation of knowledge. The future of AI is now more open and collaborative than many ever envisioned, and economic incentives dictate that innovators pro-actively capitalise on these qualities rather than trying to suppress them.

5. Interconnections among AI, cybersecurity, climate resilience, and healthcare

The investigation of the interactions between artificial intelligence, cybersecurity, climate resilience and healthcare finds synergies and threats. Cybersecurity inventions that use AI improve prediction and reaction to the cyberattacks of the engineering infrastructure. The modeling and technology with AI is helpful to promote climate resilience, particularly adaptation. AI are beneficial to healthcare delivery and the sphere of public health, although concerns of privacy, efficacy, and equity should have the first priority. Common data standards are the key to cross-sectoral interoperability and data governance.

AI will influence all the spheres of economy. There is a proliferation of AI applications in the transportation, physical sciences, energy, social services and health care. Such technologies do have a real potential in improving climate change and other disaster resiliency. The use of AI in the field of cyber applications would be enough to occupy an entire volume.

Regardless of the distinct synergies, there are significant threats that tend to jeopardize the safety, availability and resilience of AI systems in these sectors. Attacks on AI systems in the deception of cybersecurity lists could be disastrous, as it was in the case of older terrorist attacks of the simpler systems.. Defensive measures must therefore go beyond the traditional approach of securing individual systems and networks and move toward an enterprise-wide capability that protects all information-dependent components against all potential vulnerabilities [4,34-36]. Perhaps the greatest operating surprise of the COVID-19 pandemic has been the impact of supply chain disruptions on economy-wide functioning.

5.1. AI-enabled cybersecurity and threat intelligence

Integrating artificial intelligence, cybersecurity, and threat intelligence synergistically shapes an AI-secure United States. Cybersecurity has reached a tipping point, with models detecting three-quarters of new malware variants before antivirus signatures are released. AI models detecting websites distributing ransomware are already deployed, and others identify fake links. Security dashboards correlate

cybersecurity model outputs with the broader threat landscape. However, much remains to be done; new classes of threats require AI models as central components, models must migrate from security operations centers to desktop tools, and threat models need augmentation with intelligence from deep learning.

AI is poised to improve the cybersecurity of infrastructure on several fronts: monitoring for integrity violations; providing resiliency; and augmenting state monitoring to identify potential compromise of services. AI makes it possible to model large, distributed systems that have nonlinear dynamics that embodies valuable details on reliability and resiliency. AI-based models of vulnerabilities in construction can specify safety nets by detecting alternative, less-feasible transit paths or redirecting resources to the repair process in a short amount of time. AI models advise crisis and recovery management on the high-priority routes of transit and track any cross-border or regional change (e.g., radical conflict in the Russian Federation or natural catastrophe in the Caribbean) that needs a more significant resource mobilization.

5.2. Climate resilience through AI-driven modeling and adaptation

Existential hazards to humanity have been global warming. The AI may help in developing robust solutions through improvement of climate models, environmental senses, and warning metrics but the degree of value is in need of improved knowledge on the extent of benefits and how uncertainty affects expenses and rates of adjustment. In addition, the value of climate resilience investment is not entirely grounded and fails to legitimize the investment in terms of the likelihood of co-benefit with other priorities in the policy, such as security or economy.

The provision of resilience of climate effects depends on the consideration of the risks associated with climate change in planning and fast implementation of various strategies aimed at minimizing the exposure of the natural systems and communities and the economy to climate changes. One of them is the area of modeling and adaptation, which examines climate-disruption hazard by computational models with adaptive feedback to establish technical projects (e.g. tide barriers, infrastructure site, site or upgrading location and management of forests and agricultural land) and behavioral and policy adaptation choices (e.g. mobility, resettlement, voluntary buy-out, prudent water resources management, forest and agricultural development and management) which minimize exposure to extreme events of climatic conditions.

Disruption Lieberman et al. (2022) simulate the probability of flooding of New York City due to the storm tide of hurricane storms (with consideration of the rising water table and uncertain values of probabilistic vigor of the excess water as a way of water). The findings indicate that flood bars would be cost-efficient following prosaic planning, but this would be overwhelmed by the investment of avoided and managed retreat [37-40]. The risk-benefit analysis of physical investments, as is being practiced where the asset-response time constant is low relative to timescale of the phenomenon, is less possible where avoidance or control of withdrawals is the primary option. The relative costs and uncertainties of modeling a particular risk versus the benefits of adapting other-than-modeling may inform decisions on avoiding or retarding the physiological movement of decision-making climate models.

5.3. AI applications in healthcare delivery and public health

Artificial intelligence has enormous possibilities towards quickening solutions in healthcare provision and health in the general population. The current advancement of the biological sciences and AI, unprecedented in parallel, is an opportunity of unusual characteristics to improve human health and well-

being, which should be managed carefully to achieve as many positive as negative outcomes as possible. AI duality in regard to such uses, where it can be used to improve human health, and also to create a threat, be it the misuse of personal information, by either well-intentioned or malicious agents, requires objective assessment of whether there can be a major gain through collective action taken in a well-considered manner. This must be accompanied by caution so that AI products that result are elaborate, functional, and fair enough.

An industry discussion is suggested to find out the applications of quantitative AI capabilities and dimensions in U.S. public health. Comprising the bottom-up-built vision, which was based on domain experts of public health, epidemiology, social epidemiology, AIA-related research, and data-privacy research, the resulting vision linked the emphasis of the AI-aided capabilities in all health-and-social-related field with the requirements that are essential to successful implementation: strong epidemiological evidence, resistance to sloppy design and behavior, equity in the process of creating, deploying, and using the system. The similarities between health delivery and public health utilization of AI-enabled capabilities should be the pillars of a consistent two-sector roadmap to investing in, co-evolving and co-using. The spreading intelligence in such capabilities will create enhancing returns in both health delivery and population health sectors, and the interoperability across sectors is essential in acquiring the cross-sector data, protocol, and model sharing through a common set of data standards, protocols, and models.

5.4. Cross-sector interoperability and data governance

Cross-sectoral interoperability lies in the nexus of three urging security issues, namely cyber-security, climate resilience, and public health. AI-controlled cybersecurity targets the identification and shielding of security threats on interrelated networks and computerized gadgets such as the Internet of Things. In that regard, government agencies, firms, and academia are developing AI-enabled models and dashboards that combine information on cybersecurity offered by various sources, extract patterns and trends, and generate actionable insights and predictions to the government agencies and the commercial sector. Better threat knowledge and quicker response to such incidents will, subsequently, enhance climate igneousness and empower a cold-blooded and more efficient reaction to pandemic and biological games. Such endeavors require the assistance of AI-based models and tools that would combine data on several sources and reflect the interactions between climate variables and make the future projections less ambiguous.

Another field of AI in the future is the sphere of public health, both in terms of optimization of the delivery of healthcare services and in the identification and prevention of the occurrence of biological threats. Rampant use of AI application in predictive maintenance, disease prevention, and early warning must be supported by a proper framework that guarantees privacy maintenance in addition to that the analysis has excellent validation, as well as, systems should be fair and work to all classes of the society. Cross-sector interoperability and data governance will be used in all of these areas to buttress the idea that the advantages of the development and launch of these AI applications will be achieved. It is anticipated to establish standard data formats of critical assets and specifications of data sharing and governance protocols that can be reflective of the successful community biographies that are captured in the principles of digital public goods [4,41,42]. This will avoid any solution possible by identified and established repositories by the government and allow the fast and free deployment of AI-powered solutions, diagnosis, and advice across all industries.

6. The path to U.S. leadership in AI-driven security

To achieve AI security as a nation, the United States must lead the development of AI technologies and capabilities with direct application to the security of the country and its allies—without causing unwanted divisions, antagonism, or friction. This kind of leadership involves making investments that are hard to replicate such that there is shortage of resources and the development of collaborative structures that form the basis of interdependencies with no obligatory liabilities. The creation of AI-resistant cybersecurity should become the beginning of this greater strategic investment program augmented by the intensification of competencies in the highest risk fields. A short period of time is allowed to attain these purposes in the coming few years.

Since the impact of the AI echoes the various industries and fields, i.e., they generate just as much risk as opportunity, the investments should be made on a cross-sector basis without forgetting about the needs of the respective sectors. More so, effective investments in AI security should strengthen resilience in face of climate change, enhance national public health and health care, and they should be safe in use of AI. Organizing the fundamental amalgam of cybersecurity and threat-intelligence function along these cross-sector routes will represent the nearest urgency. Incentivizing regional data connectivity and quality and diversity of AI-analysis training data is essential if U.S. security services are to preserve a leading edge in threat modeling, extreme-activity anticipation, scaling of data-driven-shared-threat-intel-to-action, and rapid AI-cyber-and-physical incident response.

6.1. Strategic investment priorities

America's investment priorities for AI-driven security must address gaps in key national and economic security capabilities that rely heavily on AI foundation models or possess significant other AI interconnections to maximize national security impact per dollar spent. In parallel with this effort, the United States must make major advances in other areas of cybersecurity, climate resilience, and healthcare delivery and public health, where there are strong interconnections with AI, focusing on objectives that are symbiotic and mutually beneficial.

The U.S. government must prioritize investments that catalyze these desired outcomes and meet internally derived deadlines—particularly the pillars related to AI and cybersecurity—if the country aspires to achieve and maintain global leadership in AI-driven security. In this regard, the following timetable is useful: an integrated AI, cyber, and incident response dashboard should be established with sufficient supporting infrastructure within two years; AI-enabled modeling for climate adaptation and investment decision support that significantly reduces the current range of uncertainty must be completed within three years; and key aspects of AI-related socioeconomic inequality must be addressed within four years.

6.2. Metrics, accountability, and independent evaluation

To create a balance between the grand visions and being a responsible steward, proper metrics, accountability, and third-party assessment should be employed. Short-term strategic investments will provide the background to a unified whole-of-nation strategy that guarantees a well-thought road-map to AI-enhanced safety, security and prosperity. The leaders would have to be rational about demand, anticipate the unintended consequences, allocate their resources wisely, and be held accountable to the people of America in order to address risk and at the same time allow innovation to take place.

Nor or so Should AI make us secures To secure all by rational aims? The response has an excellent ambition statement in a graceful order of priority: A secure America. It comes down to measurement As a

design and construction methodology means that independent metric dashboards must clearly indicate edges of work being either too slow or too fast, and the system of giving early warnings on what can be done to keep changing investments in accordance with progress. These dashboards have as well the ability of offering congressional control over funding advancement within capability gaps. The system of metrics, namely, all stakeholder-facing dashboards, must encompass the ones concerning the partnerships between the government and companies in terms of talent development, the transformation of AI-powered disaster management, and the disproportional shift of the economy and society into the safe and secure future of AI. The candidate dashboards should also be useful in streamlining and calibrating the federal capital and research and development grants to very small and new companies, and even direct the financing of self-governing science.

6.3. International leadership and coalition-building

To establish an AI-Secure America, the international cooperation on all levels is necessary, as the formulation of better relations between governmental services near the house, or even more sophisticated trade agreements, the big plan in the U.S. leadership is to be integrated into a vivid multigovernment strategy. In ten years or fewer, the officials believe that able AI-supported tools and methods will be available to implement in cybersecurity, climatic research and response, medical care delivery and communal health, and economic and military resistance. In order to realize these dreams, the United States needs to go first by investing in excellence, embracing the transformative power of AI, and rapidly deploying solutions in vehement partners all over the world. Advancements herein will have a strong input in AI-assisted national security and diplomacy, in addition to economic gain.

With the availability of AI-driven cybersecurity, climate resilience, healthcare provision, and economic-military resilience, the United States will have a chance to show the sagacity and efficacy of its actions.. The nation’s affiliates—either on an individual basis or in coalitions—should therefore be prepared to take specific actions as their solutions mature. The form of these actions, including key demonstrations and relevant policy changes, will depend on time-sensitive factors. Nonetheless, an ambition statement, articulating what success would mean and how fast it might happen, would help to keep those factors in view. Such statements should take a long-term and holistic view, covering all sectors of society and all aspects of the risk equation. International coalition-building across multiple sectors of society—firewalls, frameworks, and models for capability development, deployment, and demonstration—also offer promise.

7. Conclusion

The future of America as an open, democratic lifestyle is determined by becoming AI-secure, on leveraging the paradigm-shifting power of artificial intelligence as a marketplace to improve national security, economic security, and the social contract. To be AI-secure, the United States must develop the systems and competencies needed to make the country both fundamentally more secure against all manner of risks and satisfyingly secure for all Americans against disruption, denial, and dislocation. Becoming AI-secure entails, directional investments successfully leveraging the concept of AI throughout the economy to achieve desired outcomes within articulated timeframes. It also needs to cross policy boundaries and ensure interconnectivity of the policy domains- education, health care, climate risk, and cybersecurity- and ensure that the development in one field helps improve the others. Most importantly, it requires leadership- not compromising but firm leadership that will make the nation pull together around a common cause during which allies and partners will be welcomed on board in a working search of a common solution.

In order to achieve this dream, one has to make investment decisions. The exploitation of the opportunities and the alleviation of the risks provide the grounds to become AI-secure. The vision is elaborated with seven investment priorities, recommendations of timelines and mechanisms to oversee and evaluate it. These concerted efforts denote the course of creating American leadership, in contradiction to obvious interests-based parameters, in deploying AI as the impetus of security within the realms of cyberspace, economy, society, and environment. What is then needed is a specific implementation plan that establishes responsibility, accountability, and deadlines across the full range of departments and agencies with relevant mandates.

References:

[1] Wang H, Fu T, Du Y, Gao W, Huang K, Liu Z, Chandak P, Liu S, Van Katwyk P, Deac A, Anandkumar A. Scientific discovery in the age of artificial intelligence. Nature. 2023 Aug 3;620(7972):47-60.

[2] Waisberg E, Ong J, Kamran SA, Masalkhi M, Paladugu P, Zaman N, Lee AG, Tavakkoli A. Generative artificial intelligence in ophthalmology. Survey of ophthalmology. 2025 Jan 1;70(1):1-1.

[3] Panda S. Scalable Artificial Intelligence Systems: Cloud-Native, Edge-AI, MLOps, and Governance for Real-World Deployment. Deep Science Publishing; 2025 Jul 28.

[4] Verganti R, Vendraminelli L, Iansiti M. Innovation and design in the age of artificial intelligence. Journal of product innovation management. 2020 May;37(3):212-27.

[5] Swain P. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. Deep Science Publishing; 2025 Aug 6.

[6] Shivadekar S. Artificial Intelligence for Cognitive Systems: Deep Learning, Neuro-symbolic Integration, and Human-Centric Intelligence. Deep Science Publishing; 2025 Jun 30.

[7] Shivadekar S, Halem M, Yeah Y, Vibhute S. Edge AI cosmos blockchain distributed network for precise ablh detection. Multimedia tools and applications. 2024 Aug;83(27):69083-109.

[8] Shimizu H, Nakayama KI. Artificial intelligence in oncology. Cancer science. 2020 May;111(5):1452-60.

[9] Schwendicke FA, Samek W, Krois J. Artificial intelligence in dentistry: chances and challenges. Journal of dental research. 2020 Jul;99(7):769-74.

[10] Rashidi HH, Pantanowitz J, Hanna MG, Tafti AP, Sanghani P, Buchinsky A, Fennell B, Deebajah M, Wheeler S, Pearce T, Abukhiran I. Introduction to artificial intelligence and machine learning in pathology and medicine: generative and nongenerative artificial intelligence basics. Modern Pathology. 2025 Apr 1;38(4):100688.

[11] Swain P. Challenges and opportunities in modern artificial intelligence systems: A focus on natural language processing. The Artificial Intelligence and Machine Learning Blueprint: Foundations, Frameworks, and Real-World Applications. 2025:46-67.

[12] Panda SP, Padhy A. Business Intelligence with Power BI and Tableau: Cloud-Based Data Warehousing, Predictive Analytics, and Artificial Intelligence-Driven Decision Support. Deep Science Publishing; 2025 Aug 15.

[13] Shivadekar S, Kataria DB, Hundekar S, Wanjale K, Balpande VP, Suryawanshi R. Deep learning based image classification of lungs radiography for detecting covid-19 using a deep cnn and resnet 50. International Journal of Intelligent Systems and Applications in Engineering. 2023;11:241-50.

[14] Padhy A. Artificial Intelligence-Driven DevOps: Automating, Optimizing, and Securing Modern Software Delivery. Deep Science Publishing; 2025 Aug 26..

[15] Muppala M. Artificial intelligence, IoT, and sensor technologies for Marine monitoring and climate resilience. Digital Oceans: Artificial Intelligence, IoT, and Sensor Technologies for Marine Monitoring and Climate Resilience| Deep Science Publishing. 2025 Jul 8.

[16] Novelli C, Taddeo M, Floridi L. Accountability in artificial intelligence: What it is and how it works. Ai & Society. 2024 Aug;39(4):1871-82.

[17] Nguyen P, Shivadekar S, Chukkapalli SS, Halem M. Satellite data fusion of multiple observed XCO2 using compressive sensing and deep learning. InIGARSS 2020-2020 IEEE International Geoscience and Remote Sensing Symposium 2020 Sep 26 (pp. 2073-2076). IEEE.
[18] Mohapatra PS. Artificial Intelligence-Powered Software Testing: Challenges, Ethics, and Future Directions. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:163.
[19] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[20] Muppala M. SQL Database Mastery: Relational Architectures, Optimization Techniques, and Cloud-Based Applications. Deep Science Publishing; 2025 Jul 27.
[21] Shivadekar S. Cognitive Artificial Intelligence for Health and Climate: Deep Models, Interpretability, and Decision Support. Deep Science Publishing; 2025 Aug 4.
[22] Mohapatra PS. Artificial intelligence and machine learning for test engineers: concepts in software quality assurance. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:17.
[23] Liu SY. Artificial intelligence (AI) in agriculture. IT professional. 2020 May 21;22(3):14-5.
[24] Law R, Ye H, Lei SS. Ethical artificial intelligence (AI): principles and practices. International Journal of Contemporary Hospitality Management. 2025 Jan 2;37(1):279-95.
[25] Kshetri N, Dwivedi YK, Davenport TH, Panteli N. Generative artificial intelligence in marketing: Applications, opportunities, challenges, and research agenda. International Journal of Information Management. 2024 Apr 1;75:102716.
[26] Jeysudha A, Muthukutty L, Krishnan A, Shivadekar S. Real Time Video Copy Detection using Hadoop. International Journal of Computer Applications. 2017;162(9):42-5.
[27] Hanna MG, Pantanowitz L, Dash R, Harrison JH, Deebajah M, Pantanowitz J, Rashidi HH. Future of artificial intelligence (AI)-machine learning (ML) trends in pathology and medicine. Modern Pathology. 2025 Jan 4:100705.
[28] Panda SP, Muppala M, Koneti SB. The Contribution of AI in Climate Modeling and Sustainable Decision-Making. Available at SSRN 5283619. 2025 Jun 1.
[29] Shivadekar S, Mangalagiri J, Nguyen P, Chapman D, Halem M, Gite R. An intelligent parallel distributed streaming framework for near real-time science sensors and high-resolution medical images. In50th International conference on parallel processing workshop 2021 Aug 9 (pp. 1-9).
[30] Mohapatra PS. Artificial Intelligence-Driven Test Case Generation in Software Development. Intelligent Assurance: Artificial Intelligence-Powered Software Testing in the Modern Development Lifecycle. 2025 Jul 27;4:38.
[31] Giuggioli G, Pellegrini MM. Artificial intelligence as an enabler for entrepreneurs: a systematic literature review and an agenda for future research. International Journal of Entrepreneurial Behavior & Research. 2023 May 4;29(4):816-37.
[32] Cukurova M. The interplay of learning, analytics and artificial intelligence in education: A vision for hybrid intelligence. British Journal of Educational Technology. 2025 Mar;56(2):469-88.
[33] Chen E, Prakash S, Janapa Reddi V, Kim D, Rajpurkar P. A framework for integrating artificial intelligence for clinical care with continuous therapeutic monitoring. Nature Biomedical Engineering. 2025 Apr;9(4):445-54.
[34] Bidyalakshmi T, Jyoti B, Mansuri SM, Srivastava A, Mohapatra D, Kalnar YB, Narsaiah K, Indore N. Application of artificial intelligence in food processing: Current status and future prospects. Food Engineering Reviews. 2025 Mar;17(1):27-54.
[35] Bankins S, Formosa P. The ethical implications of artificial intelligence (AI) for meaningful work. Journal of Business Ethics. 2023 Jul;185(4):725-40.
[36] Banh L, Strobel G. Generative artificial intelligence. Electronic Markets. 2023 Dec;33(1):63.

[37] Alqahtani T, Badreldin HA, Alrashed M, Alshaya AI, Alghamdi SS, Bin Saleh K, Alowais SA, Alshaya OA, Rahman I, Al Yami MS, Albekairy AM. The emergent role of artificial intelligence, natural learning processing, and large language models in higher education and research. Research in social and administrative pharmacy. 2023 Aug 1;19(8):1236-42.
[38] Panda SP, Koneti SB, Muppala M. Benefits of Site Reliability Engineering (SRE) in Modern Technology Environments. Available at SSRN 5285768. 2025 May 1.
[39] Ahmed I, Jeon G, Piccialli F. From artificial intelligence to explainable artificial intelligence in industry 4.0: a survey on what, how, and where. IEEE transactions on industrial informatics. 2022 Jan 27;18(8):5031-42.
[40] Ahmad SF, Han H, Alam MM, Rehmat M, Irshad M, Arraño-Muñoz M, Ariza-Montes A. Impact of artificial intelligence on human loss in decision making, laziness and safety in education. Humanities and Social Sciences Communications. 2023 Jun 9;10(1):1-4.
[41] Agha RA, Mathew G, Rashid R, Kerwan A, Al-Jabir A, Sohrabi C, Franchi T, Nicola M, Agha M. Transparency in the reporting of artificial intelligence–the TITAN guideline. Premier Journal of Science. 2025;10:100082.
[42] Panda S. Artificial Intelligence for DevOps and Site Reliability Engineering: Theories, Applications, and Future Directions. Deep Science Publishing; 2025 Aug 7.