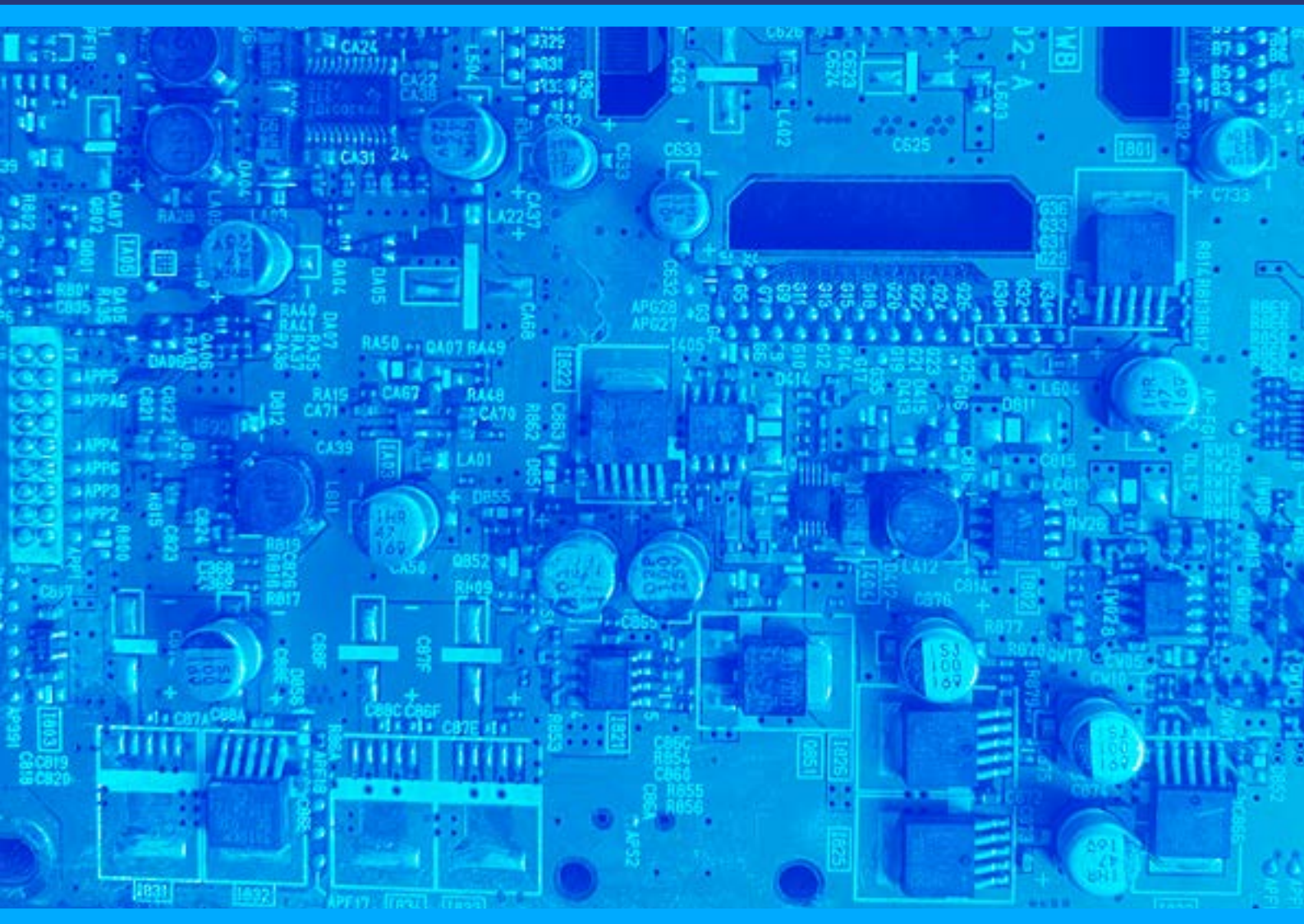
AI: Implications for Peace and Security

Articles from the SYP Conference 2024



Edited by Orlanda Gill and Tim Street

Student / Young
Pugwash UK

CONTENTS

INTRODUCTION

Student / Young Pugwash (SYP) organises an annual conference on peace and disarmament. The subject of the 7th conference, held in 2024 at King's College London, was 'Artificial Intelligence: Implications for Peace and Security'.Our aim was to encourage new thinking on the legal, political and technical questions associated with this topic, with a focus on ethical science. The articles in this collection were written by some of those who presented at the conference. The authors cover a range of topical and important ground concerning the past, present and future of AI, including in relation to nuclear energy and nuclear weapons.

• A review of the 2024 conference, with videos of each session is available here:

https://britishpugwash.org/review-of-the-7th-annual-syp-conference-ai-peace-and-security/

• A subsequent webinar on AI and nuclear matters is available here:

https://britishpugwash.org/syp-uk-webinar-ai-and-nuclear-matters-29-03-24/

• SYP supports students and young people to take part in debates concerning emerging and disruptive technology. If you would like to get involved with SYP's work, whether writing for us or taking part in events, contact Tim Street, SYP Coordinator, email: syp@britishpugwash.org

Membership of SYP is free for students and under-30s.

To learn more or get involved with SYP, visit our website:

https://britishpugwash.org/student-young-pugwash/get-involved/

Artificial Intelligence for the Peaceful Use of Nuclear Energy: Future Prospects

Syeda Saba Batool
Research Officer at Strategic Vision Institute (SVI) / Former Teaching and Research Assistant & MPhil International Relations at School of Politics and International Relations, Quaid i Azam University Islamabad / sababatool72@gmail.com

The rapid advancement of Artificial Intelligence (AI) creates a transformative opportunity for the peaceful use of nuclear energy. Over the last decade, AI has demonstrated its ability to address complex challenges in a variety of industries. AI techniques in nuclear science, such as experiment design, advanced data analysis, and theoretical modelling have aided research and development efforts. These advancements are especially visible in fusion research, where AI-powered simulations and modelling have hastened scientific progress.

Furthermore, AI plays an important role in improving the efficiency, safety, and sustainability of nuclear power plants. AI optimisation techniques improve performance, while lowering maintenance costs, by enhancing operations and reactor design. Machine learning algorithms enable real-time monitoring, predictive maintenance, and anomaly detection to ensure consistent energy production. Additionally, AI enhances nuclear security and radiation protection by analysing data from radiation detection systems and detecting intrusions in nuclear facilities.

Regulatory Frameworks for AI Applications

Despite the promising prospects for AI in the nuclear sector, several challenges remain. Regulatory frameworks must evolve to ensure responsible AI applications in nuclear material production. An AI-proficient workforce is also required to fully realise the potential of AI technologies. By addressing these challenges and promoting responsible governance, industry collaboration, and ethical standards, AI-driven innovations can contribute significantly to global prosperity and security.

Future Prospects: Expanding Horizons for AI in Nuclear Energy

The future of AI in nuclear energy holds enormous promise, with ongoing advancements poised to transform various aspects of the industry. As researchers and practitioners continue to investigate the synergies between AI technologies and nuclear applications, several possibilities emerge. AI-driven predictive maintenance systems in nuclear power plants can help to improve safety and reliability. By analysing massive amounts of operational data in real time, AI algorithms can detect potential problems before they escalate, improving safety and reliability.

The Role of AI Robotics

In one incident, a worker inadvertently fell into the nuclear spent fuel pool during inspection activities. To avoid such happenings and enhance safety protocols for workers and nuclear safeguard personnel, AI robotics are now deployed at a few facilities within nuclear spent fuel pools for inspection and related functions. Advancements in AI robotics and automation will allow for greater autonomy in nuclear facilities. From routine inspections to emergency response scenarios, AI-powered robots can complete tasks with precision and efficiency, reducing human intervention and occupational hazards. AI optimisation techniques will continue to improve nuclear processes by enhancing reactor performance, fuel utilisation, and waste management. Operators can increase the efficiency and cost-effectiveness of nuclear energy production by utilising predictive analytics and machine learning algorithms. AI robots are also being used in a few facilities, replacing human labour, yet saving workers from harmful radiation.

Nuclear Research, Safety and Security

Moreover, AI-powered simulations and modelling have the potential to accelerate advances in fusion energy research. By simulating complex plasma interactions and reactor designs, AI algorithms can help scientists achieve long-term fusion reactions,

unlocking a virtually limitless source of clean energy. AI-powered security systems will also play an important role in protecting nuclear facilities and materials from potential threats. Advanced surveillance and anomaly detection algorithms can improve perimeter security and assist in detecting suspicious activities, thereby strengthening nuclear security measures. AI technologies will drive innovation in nuclear waste management, allowing for the development of more efficient recycling and disposal methods. From isotopic analysis to waste classification, AI algorithms can optimise waste treatment processes, lowering environmental impact and long-term storage costs.

AI for Civilian Nuclear Applications

The International Atomic Energy Agency (IAEA) recognises the importance of AI in advancing nuclear applications and has launched a number of projects to encourage collaboration in this area. The IAEA works on initiatives like the AI for Atoms platform to promote knowledge sharing, collaboration and capacity building for AI applications in nuclear energy. Such platforms will accelerate the global adoption of AI-driven solutions by allowing researchers, policymakers, and industry stakeholders to collaborate better.

The future prospects for AI in nuclear energy are characterised by transformative advancements across safety, efficiency, security, and innovation domains. By harnessing the power of AI technologies, the nuclear industry can overcome existing challenges and unlock new opportunities for sustainable energy production and global development. As research and development efforts continue to evolve, AI-driven innovations will play a pivotal role in shaping the future of nuclear energy for generations to come.

Challenges to AI Applications for Advancing Civilian Nuclear Projects

There are a few challenges to utilising AI for peaceful applications of nuclear energy, including: ensuring the safety and security of nuclear facilities, managing complex datasets, and addressing ethical concerns. Moreover, balancing the imperative for accurate decision-making algorithms with the inherent unpredictability of nuclear systems, alongside the necessity of transparency and accountability, presents significant hurdles to harnessing AI for this purpose.

Policy Recommendations

The integration of AI into the nuclear energy sector, poses the challenge to policymakers of how to effectively deploy these technologies while maintaining accountability. Several policy options exist to address this issue responsibly:

• Policymakers can establish comprehensive regulatory frameworks tailored to govern AI's use in nuclear energy applications. These frameworks should encompass safety, security, ethical considerations, data privacy, and liability issues associated with AI-driven systems in nuclear facilities.

• Allocating funding and resources to support research and development initiatives focused on AI applications in nuclear energy is crucial. Collaboration between government agencies, research institutions, and private sector entities can accelerate innovation and technology transfer in this field.

• Investing in training and educational initiatives is critical for developing, implementing, and managing AI technologies in the nuclear industry. Specialised training in AI ethics, cybersecurity, and nuclear safety ensures that personnel are prepared to handle AI-powered systems in a safe and responsible manner.

• Another important policy option is to encourage international cooperation and information sharing on AI best practices, standards, and regulations within the nuclear energy community. Collaboration with international organisations, such as the IAEA, can help to establish global guidelines and standards for AI deployment in nuclear facilities.

• It is critical to conduct thorough assessments of the ethical, societal, and environmental implications of incorporating AI into nuclear energy. Engaging stakeholders, including civil society organisations and the public, in transparent dialogue addresses

concerns related to AI bias, accountability, and unintended consequences.

• Developing risk mitigation strategies for potential cybersecurity threats and vulnerabilities associated with AI-enabled nuclear systems is critical. Implementing strong cybersecurity measures such as encryption protocols, intrusion detection systems, and access controls protects against cyberattacks and data breaches.

• Establishing mechanisms for continuous monitoring and evaluation of AI deployment in nuclear energy facilities is vital. Regular audits and assessments ensure regulatory compliance, the identification of emerging risks, and promote continuous improvement in AI governance and management practices.

• Drafting regulatory frameworks to ultilize AI applications for nuclear facilities is also necessary to advance nuclear safety and security culture in AI domain.

Conclusion

The potential of AI to enhance the peaceful use of nuclear energy is very promising, presenting transformative opportunities for the industry. Despite this, several challenges persist, including the need for evolving regulatory frameworks and the cultivation of an AI-proficient workforce. By addressing these challenges and promoting responsible governance, industry collaboration, and ethical standards, AI-driven innovations can significantly contribute to advancing peaceful nuclear energy production. Policymakers will play a pivotal role in fostering the responsible and sustainable integration of AI technologies into the nuclear energy sector, unlocking its full potential for global prosperity and security.

Decoding AI Hype: The Gap Between Expectation and Reality

Océane Van Geluwe EI&C Nuclear Safety Qualification, Business France, AKKODIS Belgium, ovangeluwe@gmail.com *The views in this article are solely the author's and not those of her employers

Few topics in technological innovation garner as much attention and speculation as Artificial Intelligence (AI). AI has emerged as both a beacon of innovation and a topic of intense security concerns. Promising to revolutionise industries, streamline processes, and enhance human lives, what we commonly refer to as AI has captured the imagination of enthusiasts and sceptics alike. Yet, beneath the glossy veneer of promises lies a complex reality: the phenomenon of AI overhype.

Overhype involves promoting or publicising an object excessively. It is difficult to remember life before November 2022, when Chat GPT was launched, or before it started to boom as a day-to-day software. As headlines tout the transformative potential of AI, from driverless cars to personalised healthcare or autonomous weapon systems, questions persist about the chasm between expectation and reality. While AI undoubtedly holds tremendous promise, the hype surrounding its capabilities often outpaces its current practical applications. This dissonance has fueled a growing discourse on the need for a more nuanced understanding of AI's limitations and potential pitfalls.

From Silicon Valley boardrooms to academic symposiums, the conversation surrounding AI has gained momentum, prompting introspection within the tech community and beyond. As stakeholders grapple with the implications of inflated expectations, a critical examination of AI's true capabilities becomes imperative.

AI: the Etymology of the Term

The term "Artificial Intelligence" was coined in 1956 during a seminal conference at Dartmouth College in Hanover, New Hampshire, USA. This conference is often called the "Dartmouth Conference" and is considered the birthplace of AI as an academic discipline.

John McCarthy, a computer scientist, who is regarded as one of the founders of AI, proposed the term. McCarthy, Marvin Minsky, Nathaniel Rochester, and Claude Shannon had gathered at Dartmouth to explore whether machines could be programmed to exhibit intelligent behaviour comparable to humans.

McCarthy and his colleagues chose the term "Artificial Intelligence" to describe this field of research, intending to convey the idea of creating machines that could mimic or simulate human-like intelligence. The term "artificial" signifies that this intelligence is man-made or created by humans. In contrast, "intelligence" refers to the ability to learn, reason, solve problems, and adapt to new situations.

Since its inception, the field of AI has evolved significantly, encompassing various subfields such as machine learning, natural language processing, computer vision, robotics, and more. However, the term "Artificial Intelligence" continues to describe the overarching goal of creating intelligent machines capable of performing tasks that typically require human intelligence.

AI Winters: is the Technology Gaining Further Momentum?

Following the conference, which failed to offer "conclusive results," there was a period commonly called the [first] "AI winter," during which the initial excitement and optimism were tamed. AI systems struggle with language understanding, perception, and computing power. This led to scepticism among funders and policymakers, resulting in decreased funding and interest in AI research.

It is worth noting that several AI winters subsequently occurred. These are periods when interest and funding for AI research significantly declined due to unmet expectations, lack of progress, or

shifts in priorities. Such winters encompass the mid-1970s, the late 1980s, or the post-Deep Learning Era (late 2010s-present). The common trait of those winters is that the technology fell victim to the hype.

The early warning signs during these periods included: overhype in the early stages, an important number of investments over a short period, rising expectations followed by failure believed to be inevitable, or a simple "perceived need to 'spread the wealth'" according to James Handler. Tools were then blamed or overlooked and cast aside for future projects. For instance, and as reported by James Handler, in 1973, the UK's Science Research Council received a report that was highly critical of the AI field, concluding that none of the discoveries made up to that point had delivered the significant impact that had been promised. Consequently, funding for AI was cut off in several universities.

These AI winters serve a dual purpose: to remind us of the challenges inherent in AI research and development—and highlight the importance of managing expectations—fostering interdisciplinary collaboration, and to ensure sustainable funding to support the continued advancement of AI technologies, while managing the general public's fears.

What can we do to Prevent the Sum of All Fears?

One can readily contend that the particular attention AI receives is partly due to the fear this technology generates. With the simple use of terms and the help of the collective and popular culture, any step in AI innovation is interpreted as the potential tipping point overthrowing humanity.

Yet the literature and the professionals warn against a misconception bias leading one to commonly imagine AI tools as the dystopic robot or computer displayed in science fiction. Some even infer that "AI does not exist", call for a reassessment of the term Artificial Intelligence, and prefer the term 'Augmented Intelligence', as the current technology does not possess a will of its own. Nevertheless, future policy must address the current dynamic, preventing fears and a potential umpteenth AI winter.

Policy Recommendations

1. The first and most obvious recommendation is that urgent efforts are needed to establish agile and adaptive regulatory frameworks that can keep pace with the rapid evolution of AI technologies. Governments and international organisations could collaborate to create frameworks that balance innovation with ethical considerations. The likelihood of success is low, as we still expect such frameworks and dialogue with information and communication technologies, i.e., the cyber domain. Therefore, the framework put in place could benefit from continuous and adaptable monitoring to address emerging risks and opportunities in the AI ecosystem, while noting the challenge of AI still having a vague definition.

2. The second regulation should encourage public-private partnerships to enhance AI education and awareness programs. This could include initiatives to upskill the workforce, disseminate accurate information about AI, and bridge the knowledge gap between policymakers and technologists. The likelihood of success would depend on stakeholder's willingness to learn and the incentive for information sharing between the public and the private sector. The third is a consequence of the previous one: the government, the private sector, and other actors should work together to provide accurate information, dispel myths, and foster informed discussions about AI, yet such a measure depends on public appetite and exposure to information channels.

3. The last recommendation is a simple test-before-deploy strategy for critical infrastructures through collaborative Regulatory Sandboxes or algorithmic impact assessments. These measures allow space to experiment with AI applications in controlled environments and introduce a mandatory assessment. This recommendation is most likely already in place in most AI producing countries, which are mostly located in the US, Western Europe, and Asia, yet it is crucial to remember the importance of taking a risk-based approach when implementing new systems.

Conclusion

The discourse surrounding AI encapsulates both boundless potential and significant challenges. As we delve into future AI development and regulation, it becomes evident that a balanced approach is paramount. We must acknowledge the lessons of history, learning from past cycles of hype and disillusionment while embracing this technology's transformative possibilities.

Effective regulation requires collaboration, agility, and a deep understanding of AI's ethical, social, and economic implications. It demands continuous monitoring and adaptation to keep pace with the rapid evolution of AI technologies and mitigate emerging risks. Education and awareness are essential pillars of responsible AI development. They empower stakeholders to engage in informed discussions and make ethical decisions. Public-private partnerships can foster AI literacy, upskill the workforce, and dispel myths and misconceptions.

Overhype and overspecialisation of the AI sphere should be prevented while allowing space for technological innovation. The challenge in regulation is to open AI up to the largest possible public engagement, so that people from all walks of life can make contributions to the field.

...

"I understood that we needed to stop just being specialists in one discipline and that it was important to listen to what others had to tell us: biologists, psychologists, sociologists, cats, and others."

L. Julia. L'intelligence artificielle n'existe pas. Editions First. p. 123. 2019.

Additional references

Grandpierre, G., "L'intelligence artificielle va-t-elle changer nos vies?", L'Union.

Winkler I and Brown, You CAN Stop Stupid: Stopping Losses from Accidental and Malicious Actions. p. 368. 2020. ISBN-10⍑:⍑1119621984

Even-Dar E. et al, Action Elimination and Stopping Conditions for the Multi-Armed Bandit and Reinforcement Learning Problems. Journal of Machine Learning Research, 2006

Xu H., et al. Robustness and Regularization of Support Vector Machines. Journal of Machine Learning Research, 2009

Mannor S. and JN Tsitsiklis. The sample complexity of exploration in the multi-armed bandit problem. Journal of Machine Learning Research, 2009

Tessler C et al. A deep hierarchical approach to lifelong learning in Minecraft - Proceedings of the AAAI conference on artificial intelligence, 2017

Zahavy T. et al. Graying the black box: Understanding dqns - International conference on machine learning, 2016

The Role of European Democratic Multilateralism in Shaping Global Military AI Governance

Mahmoud Javadi PhD researcher at the Brussels School of Governance, Vrije Universiteit Brussel (VUB), Mahmoud.Javadi@vub.be

In February 2023, during the first global Summit on Responsible Artificial Intelligence in the Military Domain (REAIM) hosted by the Netherlands, the United States unveiled the Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy, inviting states worldwide to join this effort. As of June 2024, fifty-four countries have joined this initiative, including all European Union (EU) member states.

Utilising its diplomatic and political arsenal, exemplified by the Political Declaration, the United States endeavours to harness military AI capabilities in countering its near-peer competitor, China. However, it is imperative to recognise that the approach outlined in the Political Declaration reflects a distinctly American perspective—and not a truly transatlantic or even global perspective—on regulating military AI. This divergence is notably evident in the European nations' stance, which frequently takes the position of a stalwart advocate for global norms and a regulatory powerhouse.

In this article, I argue that Europe, and the EU in particular, should combine its commitment to democratic multilateralism and its focus on strategic multilateralism to sustain momentum generated by and focused on the REAIM. In this way, Europe can create a more inclusive process, which might avoid certain downsides connected to the American approach.

The American Understanding of Military AI

In his last scholarly endeavour, Henry Kissinger, in collaboration with Graham Allison, unveiled the findings of a group of technology leaders at the forefront of the AI revolution. Their conclusion was stark: "the prospects that the unconstrained advance of AI will create catastrophic consequences for the United States and the world are so compelling that leaders in governments must act now." At the forefront of these concerns is the reckless integration of AI into nuclear command and control, potentially rendering human decision-making redundant.

Domestically, the Biden administration attempted to address this challenge in the Executive Order released on October 30, 2023, regarding the 'Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence.' Internationally, the Biden administration also expressed an aspiration to contribute to the regulation of military AI. As US Ambassador Bonnie Jenkins emphasised in her keynote remarks at REAIM, "advancements in [AI] will fundamentally alter militaries around the world, large and small." Jenkins underscored that the US approach to military AI is rooted in the principles of "safe and responsible behavior," aligning with "the law of war and international humanitarian law."

This commitment is further manifested in the strategic initiatives undertaken by the US Department of Defense, which has released two key guidelines: "AI Ethical Principles" and the "Responsible AI Strategy and Implementation Pathway," alongside its key deliverable, the "Responsible Artificial Intelligence (RAI) Toolkit," publicly released in November 2023. These strategic documents highlight Washington's dedication to fostering ethical and responsible practices in the realm of military AI. The United States argues that these steps are needed to ensure "international stability" in the global application of AI within the military domain.

Washington's overarching goal is to navigate the challenges and leverage the advantages of AI during a period of 'strategic competition,' as articulated by US Assistant Secretary of State Mallory Stewart during her address at the UN Conference on Disarmament in May 2023. Stewart's remarks tie directly

to the observation that in the current landscape, the United States and China stand as the primary AI superpowers, possessing the requisite talent, research institutions, and extensive computing capabilities necessary for training cutting-edge AI models for both civilian and military purposes.

As the 2022 National Security Strategy emphasises, the imperative for the US is to responsibly navigate the competition between Washington and Beijing. During the recent meeting between President Biden and President Xi, media reports suggested the American and Chinese leaders were poised to announce a landmark commitment to banning the use of AI in autonomous weapons, such as drones, and in nuclear warhead control. However, the official statement released after Biden's meeting with Xi did not include such a pledge. Instead, it alluded only to the necessity for bilateral government talks to address the risks associated with advanced AI systems and enhance AI safety.

Given the lukewarm support from China towards Washington's initiatives, it is strategic for Washington to focus primarily on advancing and standardising regulatory frameworks for military AI. This approach would promote and spread Washington's vision globally. Besides, the United States can strategically use the framework as a tool against China and other rivals. The Political Declaration aligns perfectly with these dual objectives.

The European Response to the Political Declaration

The Political Declaration originated during the REAIM Summit, hosted by the Netherlands. The Summit issued a declaration titled 'Call for Action,' which was endorsed by 57 states, including China. Although Ambassador Jenkins hinted in her keynote remarks during the REAIM Summit that the United States had engaged in discussions with many states about the Declaration before and during the Summit, the absence of any reference to the Declaration in REAIM's Call for Action appears to be a result of opposition from some endorsers. It may also be rooted in the dissatisfaction of Europeans, including the Netherlands, with America's co-opting of REAIM for its proposal.

However, the noteworthy characteristic of REAIM lies in its inclusive nature, a distinct feature that sets it apart from the Political Declaration, making it a pivotal aspect of the Call to Action. According to the EU statement, presented on behalf of EU member states during a UN Conference on Disarmament meeting in August 2023:

"The EU recognises the significance of promoting international cooperation and a multilateral approach to address the challenges associated with AI in the military domain. In this regard, the EU welcomes initiatives such as the first REAIM Summit, hosted in the Netherlands in February this year, and we look forward to continuation of this process with a second Summit in the Republic of Korea next year. Inclusive collaboration among States and relevant stakeholders, such as industry, civil society and academia, is essential to enhance our knowledge and understanding of this issue."

The vision for Europe's strategic autonomy and the national interests of European nations both point towards the strengthening of established coalitions and alliances, notably the transatlantic partnership. Consequently, it is in Europe's best interest to collaborate with the United States in shaping global governance on military AI. However, it is crucial that the transatlantic partnership does not compromise European endeavours to bolster multilateralism and its manifestations, such as the UN, and the search for inclusiveness.

The current disassociation of the Political Declaration from larger multilateral processes can therefore be a cause for concern among Europeans, not to mention the absence of non-state actors within the US initiative. To address these issues, proactive support from Europeans, particularly the EU, is essential to sustain the momentum of REAIM and reshape it into a catalyst for global governance on military AI. Meanwhile, it remains crucial for European values and interests to prioritise the UN as the focal point for any endeavours aimed at governing military AI.

Given that the UN serves as both the overt and

covert stage for numerous conflicts between status quo states supporting the existing liberal international order and revisionist states pushing for its rewriting, achieving a globally accepted treaty restricting the use of AI capabilities is challenging, albeit not entirely implausible. By expanding the discourse on REAIM, Europeans can contribute to fostering a common understanding of the breadth and depth of military AI governance. This, in turn, sets the groundwork for establishing a convention on military AI.

It is important for Europeans to carefully watch their steps towards this goal. Success in paving the way for military AI arms control regime essentially hinges on multilateral efforts. The key question is which types of multilateralism would work best. I argue that a blend of 'strategic multilateralism' and 'democratic multilateralism' would serve as an effective recipe for Europeans to institutionalise the REAIM discourse and emphasise the central role of the UN.

Europe perceives itself as an influential norm-setter, having played the role of a "consistent leader investing in effective multilateral solutions." However, the number of crises faced by the international liberal order and its European proponents over the past two decades raises questions about the continued validity of effective multilateralism. In light of this, the 2016 Global Strategy for the EU's Foreign and Security Policy emphasises "principled pragmatism." It encapsulates the Europeans' political philosophy of combining ambitious visions with practical steps in its external actions. The concept of principled pragmatism has driven Europe toward strategic multilateral governance. This signifies a departure from a purely apologetic stance in promoting Europe's interests. Besides, it indicates an increasing willingness to leverage economic and diplomatic resources and adopting a more problem-solving-oriented approach in selecting frameworks for cooperation.

Strategic multilateralism underscores Europe's heightened aspirations for global engagement, aiming to inspire others to emulate its approach. In contrast, democratic multilateralism embodies Europe's humility. EU High Representative Josep Borrell's statements, such as "we must listen carefully to the countries of the South" and "the world is moving in a direction that is not desired by Europe" echo Europe's commitment to a more humble role in reshaping multilateralism. Such discourses allude to democratic multilateralism. It embodies values that enhance European nations' standing worldwide. This implies that the Europe has the potential to champion an inclusive, polycentric, and fairer multilateralism based on democratic principles.

In the realm of military AI, the fusion of democratic multilateralism and strategic multilateralism empowers Europeans to collaboratively interact with diverse stakeholders, including nations, across the Global North, the Global South and the Global East. Significantly, this approach sustains robust engagement across the Atlantic while affording Europeans the flexibility to seek coalition partners globally who align with their vision for governing military AI—a vision that revolves around the REAIM objectives and places emphasis on the central role of the UN in the long run.

Dual-Use Challenges of AI in Nuclear Deterrence

Mariam Mumladze Bachelor of International Relations at the Free University of Tbilisi

Artificial Intelligence (AI) has evolved into a multifaceted technology, opening a Pandora's box of regulatory dilemmas and raising concerns about accountability, transparency, and ethical use. Its dual-use nature complicates efforts to establish clear regulatory guidelines, especially as AI is increasingly integrated into both civil and military domains.

Incorporating AI into military operations offers defensive benefits such as improving early warning systems and second-strike capabilities, and potential risks. The use of AI in unmanned platforms and the increasing focus on low-yield nuclear submarine-launched ballistic missiles (SLBMs) and cruise missiles (SLCMs) underscores a growing AI-driven arms competition among major powers such as the United States, Russia, and China. These advances heighten concerns about inadvertent nuclear escalation and emphasises the need for strong policies and robust human oversight to protect international security.

This article integrates perspectives from different sources, including a 2020 RAND report on the impact of AI on deterrence, legislative proposals such as the "Block Nuclear Launch by Autonomous Artificial Intelligence Act," and national strategies of the United States, the United Kingdom, and France. It also draws on historical analogies to illustrate the regulatory challenges posed by dual-use applications of AI in civilian and military contexts.

Global AI Military Integration

The current era of artificial intelligence parallels the historical militarisation of industrial age technologies at the turn of the 20th century such as submarines, aircraft, balloons, poison gas, and certain types of ammunition, which led to analogous efforts by states to regulate these transformative capabilities

Similar to the arms control measures before World War I and II, challenges arise today with AI as both Russia and China are actively integrating AI into their military strategies, albeit with different approaches. In authoritarian nuclear-armed states the decision to automate nuclear capabilities is influenced by regime stability, threat perceptions, and the desire to centralise control due to fears of internal instability or external threats. While Russia has been transparently focusing on AI-powered platforms, such as AI-equipped bombers, neural networks, and hypersonic vehicles, for nuclear weapon delivery, China, on the other hand, with its strict nuclear control and a centralised system, sees AI as an integral part of achieving global dominance in the military domain. China's People's Liberation Army (PLA) envisions "intelligentised warfare," considering AI necessary for strategic decision-making and transforming military operations into a scientific process based on analysis and calculation. However, democratic countries like the United States and its allies prioritise accountability, responsibility, and ethical considerations, which is why they are more cautious about integrating AI into sensitive nuclear decision-making processes.

The 2020 RAND report, for example, analyses how the spread of AI and autonomous systems could affect deterrence by stimulating possible wargame scenarios. Their results (Figure 1) show that the composition of human versus automated decision-making, as well as human versus unmanned presence significantly influences the dynamics of escalation in crises. Especially when automated machines are the main decision-makers, escalation becomes harder to control or prevent. This underscores the continued importance of human judgment in the face of potential for automation bias, as seen in the NATO Able Archer military exercise in 1983 and the accidents in 2003.

While the US, the UK, and France stress the importance of human oversight in the decision-making of their defence AI strategies, clearer guidance is urgently needed to ensure effective implementa-

Table 7.2 Human and Machine Configurations and Potential Escalatory Dynamics

Physical Presence	Decisionmaking: Primarily Human	Decisionmaking: Primarily Machine
Human	Lower escalatory dynamic Higher cost of miscalculation	Higher escalatory dynamic Higher cost of miscalculation
Machine	Lower escalatory dynamic Lower cost of miscalculation	Higher escalatory dynamic Lower cost of miscalculation

Figure 1. "Deterrence in the Age of Thinking Machines," by Wong et al, RAND Corporation, 27 January 2020, p.64.

tion. This includes establishing safeguards against AI errors and manipulation, defining the scope of human involvement in AI-enabled operations, and addressing the complex challenges of accountability in these contexts.

The United States should strongly oppose the use of "dead hand" nuclear launch systems, such as the Perimeter system which Russia reportedly uses for its nuclear arsenal, and any system with with predefined launch commands that include algorithmic elements. Although some are advocating these systems in response to new threats such as AI and hypersonic missiles, the risks they pose are too great. On the contrary, the US should focus on strengthening its ability to deter a second strike by maintaining human control over strategic systems, as proposed in legislation such as the Block Nuclear Launch by Autonomous Artificial Intelligence Act introduced by US lawmakers to prevent the use of autonomous weapons systems to launch nuclear weapons and ensure human control based on moral and strategic risks. Placing nuclear warheads on unmanned vehicles should also be rejected to secure direct human control of nuclear safeguards.

Policy Recommendations

Today the situation is complex and it is not surprising that there is a need to review and reformulate current approaches to arms control rather than following the traditional normative frameworks. The following five policy recommendations can be practical steps toward improving international security and stability:

1. Develop robust risk assessment frameworks to identify and mitigate dual-use AI technologies in nuclear deterrence. This includes multi-stakeholder consultations, ethical impact assessments, and clear human oversight protocols to prevent unintended escalation.

2. Establish clear regulations that emphasise observable behaviour in AI-driven military operations, particularly in nuclear contexts. Establish verification mechanisms such as on-site inspections and simulation exercises, modelled on the success of arms control treaties such as the SALT Treaty during the Cold War to ensure compliance with arms control agreements related to AI integration.

3. Enforce ethical guidelines for AI in nuclear deterrence. Promote transparency by requiring states to disclose AI capabilities and intentions, thereby enhancing trust and minimising the risk of misinterpretations or misjudgment.

4. Engage in collaborative dialogues, academic forums, and Track II exchanges to deepen understanding of the military implications of AI technologies, drawing on models of international cooperation from the historical discussions on the

regulation of technologies such as chemical weapons after World War I. Ensure the participation of scientists and engineers in policy discussions to base suggestions on technical reality and prepare for evolving challenges.

5. Recognise the rapid evolution and resilience of AI. Implement strategic export controls such as those applied in the Missile Technology Control Regime (MTCR), for critical AI technologies within global supply chains to minimise shipments.

Conclusion

The emergence of autonomous systems in nuclear warfare is a turning point in global political discourse. Unlike the rapid integration of nuclear weapons, the emergence of autonomous technologies allows for more targeted involvement of all sectors of society, thereby promoting collective participation and discussion.

The problem of responsibility in situations where people are sidelined requires the development of strong frameworks to assign accountability. While human-machine collaboration offers promising chances for improving warfighting capabilities, careful management is also essential to diminish the inherent risks associated with reduced human monitoring.

In addition, addressing the dual use of AI and limiting its spread reflects the challenges encountered by previous technological advances. Efforts to establish preventive norms and intergovernmental agreements are fundamental to address the risks of autonomous weapons. Although negotiations and international agreements are complex, it is essential to build consensus among countries based on common interests in preventing the spread of autonomous weapons.

How are AI start-ups revolutionising the Western defence industry? The case of Anduril and implications for Europe

Jan Quosdorf
King's College London, MA Candidate International Affairs
Hamburg University, MA Candidate Peace and Security Studies

Vincent Tadday
Sciences Po, MPP Candidate Politics and Public Policy
Hertie School, MPP Candidate Public Policy

How are Western defence industries adapting to the possibilities of applying artificial intelligence (AI) to defence purposes? As these capabilities are becoming ever more powerful and increasingly utilised, it is important to be aware of how their development is being facilitated and shaped. Already, rapidly growing tech start-ups—with an explicit focus on military applications of AI—are being established, such as US-based Anduril Industries. In the West, this is a novelty in a sector which for two decades has been dominated by a military-industrial complex featuring few established companies, and selective engagement with AI within a broader portfolio of military and civil products. To understand and contextualise this evolution, this article examines how regulatory, procurement, and geopolitical considerations intersect with national imperatives, technological advances and strategic autonomy.

Anduril's Rise in the Context of US Policies

In the past ten years, efforts to harness AI applications to defence have been prominent in US policy. Under former President Obama, hints at a more serious engagement became evident with the creation of the Defence Innovation Unit experimental (DIUx) in 2014 and a 2016 study on autonomy by the Defence Science Board. The Trump administration facilitated an expansion in this area, showcased by the publication of the 2018 Department of Defence's AI Strategy, the explicit reference to AI in the National Defence Strategy, and removal of the experimental "x" for the DIU. Under President Biden, the next step appears to be embedding AI innovation within broader efforts to stimulate the US defence industry, through policies such as the Replicator Initiative and the Data, Analytics, and AI Adoption Strategy.

This enhanced US engagement with AI is paralleled by the emergence of companies such as Anduril Industries. Led by Palmer Luckey, the venture capital company involves figures with strong ties to the Trump administration. Launched in 2017, Anduril initially worked closely with Homeland Security and the Defence Innovation Unit, with the first major project being its virtual border wall, a system of sensor towers to detect people in a specified area.

By now, Anduril's portfolio includes a variety of autonomous hardware products, which revolve around Lattice OS, its AI-powered operating platform. Lattice serves as the central command and control system enabling the connection of multiple autonomous assets. Anduril has developed such assets for land, oceans and the sky, for example the underwater and air vehicle "Dive-LD" and "Roadrunner". What differentiates Anduril from established defence companies is not their innovation within this domain, or the promotion of autonomous solutions for defence, but their almost exclusive focus on this technology.

The company made headlines in the wake of the 2018 Project Maven controversy. Initially assigned to Google, this US Air Force-led project aims at training an AI system for image analysis purposes, based on drone footage. While claiming to be intended for non-offensive uses only, public knowledge about Google's involvement led to employee protests, and statements of opposition at other Silicon Valley companies, such as Microsoft. Amidst this wider public debate, Anduril's leadership spoke out in favour of tech companies contributing to US military innovation, and was ultimately assigned the project alongside Palantir, after Google did not extend the contract.

Trends in the US Defence and Military Innovation Complex

The adoption of AI for military purposes has subsequently been shaped by governmental efforts and corresponding industry developments. Open innovation systems play a more prominent role in this domain, as the US seeks to build and capitalise on a broader innovation ecosystem beyond the state's control. Through those initiatives the US military wants to foster collaborative environments where private sector companies, and academic institutions share knowledge, data, and technology to accelerate the development and deployment of AI solutions for defence applications. These platforms facilitate innovation by leveraging a diverse pool of expertise and resources, fostering partnerships, and enabling rapid prototyping and testing of AI technologies to enhance military capabilities. As the military faces declining relative shares in terms of sales and research and development budgets in computing and semiconductor industries, it seeks to develop shorter feedback loops and institutionalised relationships with non-traditional vendors. A critical question will concern how this practice could influence the aerospace or nuclear domain in the future.

Traditional characteristics of US military innovation could also be subject to change. While different military branches have engaged with AI technology, the Air Force seems to be leading the race, as it embraces the digitalisation of the battlefield. With most US-based PhD graduates in the AI field still coming from abroad, future screening processes and de-risking policies could raise barriers to immigration and thus limit it as an innovation factor. The 2018 Google protests also exemplify how the desire to avoid casualties restricts tech companies from contributing to military innovation, and the implementation of fully autonomous systems in defence.

However, organisational changes and available financial support, including from venture capitalists and investors, have led to a wave of start-ups being created which are willing to fill the empty slot. The emergence of companies like Anduril or Shield AI could seriously challenge the post-Cold War consolidation of the US defence industry. Judging from previous declarations, it is likely that US governmental interests in stimulating innovation by promoting competition, and the perception of comprehensive regulation as an obstacle, are likely to increase if great power tensions continue to intensify.

Implications for the European Defence Market

The evolution of the US defence industry, driven by the emergence of innovative companies like Anduril Industries, has already prompted reactions within the European defence market. Europe is lagging behind the US in terms of funding for military AI research and development. Furthermore, entering the European defence market remains challenging for companies like Anduril. This difficulty stems from three main barriers.

Firstly, the regulatory and strategic landscape surrounding the military use of AI in European Union (EU) member states is fragmented. With the exception of France, few EU countries have developed comprehensive AI defence strategies. At the EU level, the recently introduced EU Artificial Intelligence Act (AI Act) explicitly excludes military AI, further exacerbating the regulatory gap. This lack of cohesive strategy and regulation hampers the integration of AI technologies into European defence systems.

Secondly, the complex procurement processes in European defence markets, often characterised by bureaucratic hurdles and decentralised decision-making, pose significant challenges for US defence start-ups. Navigating these intricate procedures requires a deep understanding of local regulations and cultural norms, as well as established relationships with key stakeholders, which US start-ups may lack. Moreover, the traditional focus on cost-effectiveness and risk aversion in European procurement can limit opportunities for innovation, as governments may be hesitant to adopt unproven technologies or vendors. This risk aversion stifles the agility and disruptive potential of US start-ups, hindering their ability to penetrate markets where there is resistance to change and a preference for

established solutions.

In addition, leading European defence companies, such as Thales, have a better understanding of the procurement process and long-established relations to decision makers. Those companies are investing heavily in AI research and integration, with initiatives like the creation of the cortAIx AI accelerator to expand AI into Thales' defence systems, which are already used by many European nations.

Thirdly, European governments prioritise domestic and regional suppliers due to concerns over sovereignty and dependence on American technology. This preference for regional suppliers reflects broader geopolitical considerations and underscores the challenges faced by non-European defence companies seeking to enter the market.

Fostering Innovation and Challenging Traditional Norms

The trajectory of AI integration into defence currently reflects a dynamic interplay between governmental policies and industry advancements, evident in the US and Europe. The emergence of new players, such as Anduril Industries, is indicative of a shift in the defence landscape, challenging traditional norms and necessitating ethical, political, and legal regulation. Understanding how great power competition, national and regional interests, and attitudes of private companies are tied together can serve as a basis for evaluating possible solutions.

Already, there are signs of progress to removing obstacles in the European sector. Initiatives such as NATO's Innovation Fund indicate a growing recognition of the importance of innovation and collaboration in defence technology. While navigating the complexities of the AI landscape, regulatory frameworks, and geopolitical dynamics, industry stakeholders and policymakers are increasingly acknowledging the need for a more open and dynamic environment that facilitates the integration of cutting-edge technologies into defence systems.

The case of Anduril Industries exemplifies the transformative impact AI start-ups are having on the Western defence industry. Anduril's focused approach to military AI, supported by significant ties to US governmental initiatives, highlights how regulatory and procurement policies can foster rapid innovation. This contrasts sharply with the fragmented and bureaucratic landscape of the European defence market, which poses significant barriers to similar advancements. Despite these challenges, there are encouraging signs that Europe is beginning to recognize the need for a more cohesive and open approach to integrating AI into defence. Initiatives like NATO's Innovation Fund and increased investment by companies like Thales indicate a shift towards fostering innovation and collaboration. The rise of companies like Anduril not only challenges traditional norms but also underscores the need for ethical, political, and legal frameworks to keep pace with technological advancements.

Student / Young Pugwash UK

britishpugwash.org