DEPOLITICISING AI WITHIN THE NPT REVIEW PROCESS

SANAA ALVIRA

Not surprisingly, there has been increased interest in Artificial Intelligence (AI) at the recent nuclear Non-Proliferation Treaty (NPT) meetings. However, discussions on AI in these meetings have so far been vague and overarching. More sensitive issues, such as AI in nuclear command and control, are likely to face resistance and even backlash as discussions on AI continue to evolve and gain traction. This paper argues that depoliticising the technical applications of AI—such as by highlighting the benefits of AI in supporting the International Atomic Energy Agency (IAEA) safeguards regime—could potentially act as a catalyst for limited cooperation within the NPT review process. It also offers recommendations for the 2026 Review Conference to initiate discussions on technically focussed issues on AI that could not only help overcome diplomatic obstacles to some extent, but also lay the groundwork for broader AI dialogues under the NPT if the political will develops in the future.

Ms **Sanaa Alvira** is a Research Associate at the Centre for Aerospace Power and Strategic Studies, New Delhi.

INTRODUCTION

Often referred to as the "cornerstone" of the international nuclear non-proliferation regime,[1] the nuclear Non-Proliferation Treaty (NPT) is the most widely adhered-to multilateral disarmament agreement, with 191 states-parties.[2] Under Article VIII, paragraph 3[3] of the NPT, states-parties are required to review the implementation of the treaty every five years, a commitment reaffirmed at the 1995[4] and 2000[5] Review Conferences. This review process involves a series of Preparatory Committee meetings held in the three years prior to each Review Conference. If necessary, a fourth Preparatory Committee may also be convened in the year of the Review Conference.[6] Review Conferences generally aim to adopt a final document by the consensus of all the states-parties. Together, these meetings provide a structured framework for evaluating past commitments and deciding on future actions. However, these meetings are increasingly being overshadowed by heightened geopolitical tensions and growing criticism of their structural imbalances and recurring deadlocks over the fulfilment of disarmament obligations. Notably, the 2022 Review Conference marked the first time in the NPT's five-decade history that two consecutive conferences (in 2015 and 2022) failed to reach agreement on a final outcome document.[7]

1. United Nations Office for Disarmament Affairs, *Treaty on the Non-Proliferation of Nuclear Weapons (NPT),* https://disarmament.unoda.org/wmd/nuclear/npt/. Accessed on March 15, 2025.
2. United Nations, *Tenth Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons (NPT): Background*, https://www.un.org/en/conferences/npt2020/background. Accessed on March 15, 2025.
3. n. 1.
4. United Nations Office for Disarmament Affairs, *1995 Review and Extension Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons*, https://disarmament.unoda.org/wmd/nuclear/npt1995/. Accessed on May 22, 2025.
5. United Nations Office for Disarmament Affairs, *2000 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons*, https://disarmament.unoda.org/wmd/nuclear/npt2000. Accessed on May 22, 2025.
6. Gaukhar Mukhatzhanova, "NPT Review Process: An Explainer," *Governing the Atom Brief* No. 4, Vienna Centre for Disarmament and Non-Proliferation (VCDNP), https://vcdnp.org/wp-content/uploads/2023/07/web_npt_review_process_07212023.pdf. Accessed on May 22, 2025.
7. "Pragmatic Steps to Reinforce the NPT on the Way Towards the 2026 Review Conference: Policy Brief," European Leadership Network, December 2023, https://www.

In 2023, the new review cycle began with an unpromising start, despite low expectations.[8] The third Preparatory Committee for the 2026 Review Conference concluded on May 9, 2025, after two weeks of deliberations in New York. The session concluded without adopting a set recommendations to the Review Conference or the draft decisions put forward by the chair of the committee on "strengthening the NPT review process."[9] The meeting agreed on more procedural matters, such as the dates, venue[10] and elected president of the Review Conference[11], as well as the procedural report.[12] Such an outcome, while disappointing, was not unexpected – the deliberations over two weeks (including a session of closed-door negotiations) once again highlighted the deep polarisation among countries as well as their inability and/or unwillingness to find common ground on matters of substance. When the Review Conference meets in New York in 2026, it will be 15 years since these NPT meetings reached consensus on an outcome document.

There are myriad issues that are addressed at the NPT meetings relating to the three pillars of the treaty: disarmament, non-proliferation and peaceful uses of nuclear energy and technology. However, there is one issue that

europeanleadershipnetwork.org/wp-content/uploads/2023/12/23_11_24_WP-Policy-Brief_online.pdf.

8. William C. Potter, "Behind the Scenes: How Not to Negotiate an Enhanced NPT Review Process," *Arms Control Today*, October 2023, https://www.armscontrol.org/act/2023-10/features/behind-scenes-how-not-negotiate-enhanced-npt-review-process.
9. International Campaign to Abolish Nuclear Weapons (ICAN), "No Agreement at Non-Proliferation Treaty PrepCom; TPNW States Point Way Forward," https://www.icanw.org/no_agreement_at_non_proliferation_treaty_prepcom_tpnw_states_point_way_forward. Accessed on May 23, 2025.
10. United Nations Office for Disarmament Affairs, "Decision on the Dates and Venue of the 2026 Review Conference," NPT/CONF.2026/PC.III/DEC.1, May 8, 2025, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/NPT_CONF.2026_PC.III_DEC..01_-_01.ADVANCE_UNEDITED_VERSION_-_Decision_on_the_dates.pdf.
11. United Nations Office for Disarmament Affairs, "Election of the President and Other Officers," NPT/CONF.2026/PC.III/DEC.2, May 8, 2025, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/NPT_CONF.2026_PC.III_DEC.02_-_02._ADVANCE_UNEDITED_VERSION_-_Election_of_the_President_and_other_officers.pdf.
12. Reaching Critical Will, "Draft Final Report of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons," NPT/CONF.2026/PC.III/CRP.11, May 8, 2025, https://reachingcriticalwill.org/images/documents/Disarmament-fora/npt/prepcom25/documents/CRP11.pdf.

While emerging technologies as a whole present a wide range of challenges for the NPT, using AI as a practical starting point to find common ground could help form the base for structured dialogues on other emerging technologies as well.

the NPT has been slow to address, but which is slowly and steadily gaining traction: emerging technologies. An important study on the future impacts of emerging technologies on international stability by the Institute for Peace Research and Security Policy at the University of Hamburg helpfully defines emerging technologies "as those technologies, scientific discoveries, and technological applications that have not yet reached maturity or are not widely in use but are anticipated to have a major—perhaps disruptive—effect on international peace and security".[13] These include AI, additive manufacturing, space-based capabilities, quantum computers, blockchain, hypersonics, among others. While emerging technologies as a whole present a wide range of challenges for the NPT, using AI as a practical starting point to find common ground could help form the base for structured dialogues on other emerging technologies as well. In fact, the terms 'AI' and 'emerging technology' are often used interchangeably in diplomatic discourse, highlighting the prominence of AI in current strategic thinking, and, hence, the value of taking a targeted approach towards it. This paper will focus specifically on AI, exploring avenues for its 'depoliticisation' as a first step towards limited cooperation, in order to help prevent political stagnation in the consensus-based NPT review process.

AI IN THE NPT

To date, the most prominent discussions on AI and its impact on the broader non-proliferation regime are taking place *outside* the NPT, whether it is in the academic and think-tank world (Track 1.5 or Track 2 dialogues), the P5,

13. Marina Favaro, Neil Reino, and Ulrich Kuhn, *Negative Multiplicity: Forecasting the Future Impact of Emerging Technologies on International Stability and Human Security* (Institute for Peace Research and Security Policy, University of Hamburg, September 2022), https://ifsh.de/file/publication/Research_Report/010/Research_Report_010.pdf.

or as a talking point in national policies. Many countries are now articulating their positions on AI in national, bilateral or even multilateral settings. For instance, the US Department of Defence, in the 2022 *Nuclear Posture Review*, formally stated that "in all cases, the United States will maintain a human 'in the loop' for all actions critical to informing and executing decisions by the President to initiate and terminate nuclear weapon employment."[14] In November 2023, over 40 countries endorsed a political declaration on responsible military use of AI, with support from a geographically diverse group of states, including not just US allies but those in Africa, Southeast Asia, and Latin America as well.[15] In March 2023, the United States attempted to incorporate discussions on AI within the P5 process and noted that "military applications of artificial intelligence could enable new kinds of weapon systems and change how states make decisions in crisis or conflict. The possible use of AI in an irresponsible manner by states to inform or support nuclear operations raises serious concerns about how AI systems might affect nuclear risks. This is something that we want to avoid. We need to manage potential challenges at the intersection between emerging technologies and nuclear risks."[16] A year later, in November 2024, former US President Joe Biden and Chinese President Xi Jinping jointly "affirmed the need to maintain human control over the decision to use

The possible use of AI in an irresponsible manner by states to inform or support nuclear operations raises serious concerns about how AI systems might affect nuclear risks. This is something that we want to avoid.

14. US Department of Defense, *2022 National Defense Strategy of the United States of America, Including the 2022 Nuclear Posture Review and the 2022 Missile Defense Review*, 2022, https://media.defense.gov/2022/Oct/27/2003103845/-1/-1/1/2022-NATIONAL-DEFENSE-STRATEGY-NPR-MDR.PDF. Accessed on May 14, 2025.
15. US Department of State, "Political Declaration on Responsible Military Use of Artificial Intelligence and Autonomy," 2025, https://www.state.gov/wp-content/uploads/2023/10/Latest-Version-Political-Declaration-on-Responsible-Military-Use-of-AI-and-Autonomy.pdf. Accessed on May 14, 2025.
16. Heather Williams, *The Nuclear Order and Emerging Technologies* (London: Centre for Science and Security Studies, King's College, February 2024), https://www.kcl.ac.uk/csss/assets/the-nuclear-order-and-emerging-technologies.pdf.

nuclear weapons… The two leaders also stressed the need to consider carefully the potential risks and develop AI technology in the military field in a prudent and responsible manner."[17]

In the NPT context, there has been a general hesitancy to bring AI onto the agenda (along with the other emerging technologies) in a more concrete way. The review process in recent years has faced a deepening polarisation[18] on several topics ranging from the perceived lack of progress on disarmament obligations by Nuclear Weapon States (NWS), extended deterrence, security assurances, hampered access to peaceful uses of nuclear technology, among others. More recent issues have also become a source of contention: the contested status and growing support for the Treaty on the Prohibition of Nuclear Weapons (TPNW), implications of the 2021 Australia, United Kingdom, United States (AUKUS) security partnership on the non-proliferation regime and naval nuclear propulsion, targeting of nuclear facilities during armed conflict, and gender issues are some examples. Against this backdrop, the inclusion of another contentious topic such as AI, to an already packed agenda, could exacerbate the polarisation among the NPT states-parties and further complicate efforts to achieve consensus. Put simply, there has just not been enough time to adequately devote to AI when there are more fundamental concerns about the overall health of the treaty and the broader non-proliferation regime itself.

This is not to say that there has been no mention of AI in the NPT—rather that these have remained surface-level, often couched in broad terms. Discussions on AI have so far have not gone beyond the military focus and 'human in the loop' conversations. More sensitive issues, such as open and consultative discussions on integration of AI into nuclear command and control systems are unlikely to gain traction in formal NPT settings, likely due to the NWS' reluctance to openly engage on matters that are perceived

17. Jarrett Renshaw and Trevor Hunnicutt, "Biden, Xi Agree That Humans, Not AI, Should Control Nuclear Arms," Reuters, November 17, 2024, https://www.reuters.com/world/biden-xi-agreed-that-humans-not-ai-should-control-nuclear-weapons-white-house-2024-11-16/.
18. Marianne Hanson, Tanya Ogilvie-White, Kawasaki Akira, and Muhadi Sugiono, "The Pulse: Reflections on the 2025 NPT PrepCom," Asia-Pacific Leadership Network, May 23, 2025, https://www.apln.network/analysis/the-pulse/12490.

to touch on national security prerogatives. Despite this, there has been a gradual increase in the appetite to begin sustained conversations about AI at the recent NPT meetings. Several countries expressed their concern over the rapid pace of AI advancement at the third Preparatory Committee held in April 2025.

For example, Costa Rica stated, "Emerging technologies such as artificial intelligence, quantum computing, and autonomous systems have profound implications for nuclear stability and security, requiring urgent integration into the governance of the non-proliferation regime."[19]

Indonesia stated, "The rise of new and emerging technologies in weapons of mass destruction systems pose unknown risks. We must address such developments, maintain meaningful and effective human control and ensure compliance with international law, including international humanitarian law."[20]

The Holy See, reflecting the broader concerns that go beyond strategic stability to include moral responsibility, highlighted that "the international community has a moral responsibility to advance disarmament, particularly in light of the rapid evolution of delivery systems and cyber technologies… The integration of cyber technologies into nuclear command, control, communications and early warning systems introduces vulnerabilities that could be exploited. These concerns are exacerbated by the incorporation of autonomous components, prompting critical ethical, humanitarian, legal and security issues."[21]

19. United Nations Office for Disarmament Affairs, *Statement by Costa Rica*, General Debate, April 28, 2025, Third Session of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/Costa_Rica_-_GD_-_English_and_Spanish.pdf
20. United Nations Office for Disarmament Affairs, *Statement by Indonesia*, General Debate, April 29, 2025, Third Session of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/Indonesia_-GD.pdf.
21. United Nations Office for Disarmament Affairs, *Statement by the Holy See*, General Debate, April 29, 2025, Third Session of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_

The African Group expressed deep concern "with any attempt to integrate artificial intelligence applications to managing nuclear weapons and explosive devices. Pending the total elimination of these weapons, states must maintain full and meaningful human control over nuclear weapons and their delivery systems."[22]

In a similar vein, Brazil asked the NPT meeting to "consider the potential impact of new and emerging technologies in the risk of nuclear conflict, such as the incorporation of artificial intelligence (AI) into nuclear command and control systems and other forms of autonomy, as well as the increasing risk of weaponisation of outer space. These developments not only can be profoundly destabilising, but they can also fuel further qualitative and quantitative expansions of existing arsenals."[23]

South Korea, recalling the Responsible AI in the Military Domain (REAIM) Summit[24] held in Seoul in September 2024, stated, "Facing the advent of emerging technologies such as AI, the NWS should maintain human control and involvement for all actions critical to informing and executing sovereign decisions concerning nuclear weapons employment. Such call was agreed upon by 63 States at the Seoul REAIM Summit held last September in Seoul and also is submitted to this Committee through the Working Paper by the Stockholm Initiative. Accordingly, we recommend that the RevCon call on all NWS to endorse and implement this call at the earliest opportunity."[25]

---

the_Eleventh_Review_Conference_(2025)/HOLY_SEE_-_General_Debate_NPT_PrepCom.pdf.

22. United Nations Office for Disarmament Affairs, *Statement of the African Group*, Cluster 1, May 1, 2025, Third Session of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/AFRICAN_GROUP_(E)_-_CLUSTER_11.pdf.
23. United Nations Office for Disarmament Affairs, *Statement by Brazil*, Cluster 1, May 2, 2025, Third Session of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/Brazil_-_Cluster_1.pdf.
24. "Responsible AI in Military Domain Summit," REAIM, https://reaim2024.kr/reaimeng/index.do. Accessed on May 14, 2025.
25. United Nations Office for Disarmament Affairs, *Statement by the Republic of Korea*, Cluster 1, May 1, 2025, Third Session of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, https://docs-

The Working Paper by the Stockholm Initiative[26], referred to by South Korea, is a comprehensive document flagging the various potential impacts of emerging technologies, including AI, on nuclear disarmament, non-proliferation, and the peaceful use of nuclear technology. The paper calls for greater awareness, dialogue and transparency in order to mitigate risks, while also exploring how these technologies could support disarmament, verification efforts and the peaceful applications of nuclear technology. This is the first time that a dedicated document has formally explored the potential impact of such technologies within the NPT framework.

It is clear that these statements reflect a growing recognition among countries on the need to address the impact of AI within the NPT. While arguments such as, "Diplomats should hesitate before trying to tackle the AI-nuclear convergence. Doing so in official, multilateral nuclear security dialogues risks being unproductive or even undermining consensus to reduce nuclear risks at a time when such consensus is desperately needed,"[27] are not without merit, and the NPT, is, and will remain, an important forum to have such conversations. AI is not a peripheral issue that can be kept on the margins, but a reality that is undoubtedly impacting all three pillars of the NPT. As both a treaty and a forum, the NPT allows both NWS and Non-Nuclear Weapon States (NNWS) to collectively examine AI's impacts not just on the nuclear field but also on the implementation of their NPT commitments in a formal setting. Moreover, the NPT review process plays an important role in setting norms: it reinforces existing principles and provides a platform for shaping new norms regarding the responsible application of AI in nuclear issues. Sidelining AI for the sake of short-term goals like consensus could

library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/Korea_-_Cluster_1.pdf.

26. Stockholm Initiative for Nuclear Disarmament, United Nations Office for Disarmament Affairs, "Navigating the Potential Impact of Emerging Technologies on Nuclear Disarmament, Arms Control, Non-Proliferation and Peaceful Uses of Nuclear Energy and Technology," NPT/CONF. 2026/PC.III/WP.35, April 25, 2025, https://docs.un.org/en/NPT/CONF.2026/PC.III/WP.35.
27. Lindsay Rand, "The Risk of Bringing AI Discussions Into High-Level Nuclear Dialogues," Carnegie Endowment for International Peace, August 19, 2024, https://carnegieendowment.org/posts/2024/08/ai-nuclear-dialogue-risks-npt?lang=en.

Formally institutionalising AI as an agenda item within the NPT review process must be approached carefully. It is likely that the NWS will push back at any attempts to discuss in detail the military or sensitive applications of AI within the NPT, despite being raised by the NNWS.

risk making the NPT increasingly out of touch with the technological developments that are shaping nuclear risks and responsibilities.[28] As Dr Heather Williams states, "The challenge, therefore, is how to incorporate emerging technologies into the NPT, without overloading the states' agendas, exacerbating distrust, and adding pressure – both in terms of capacity and expectations."[29]

DEPOLITICISING AI

Formally institutionalising AI as an agenda item within the NPT review process must be approached carefully. It is likely that the NWS will push back at any attempts to discuss in detail the military or sensitive applications of AI within the NPT, despite being raised by the NNWS. For example, in the third NPT Preparatory Committee in 2025, China opposed the phrase "meaningful human control" over nuclear weapons and their delivery systems in the draft document containing recommendations, stating that "the relevant language is poorly defined and there is no internationally agreed definition."[30] The revised recommendations mention the following regarding such technologies:

> Note the need to increase awareness of the risks as well as opportunities associated with the use of emerging technologies, especially in assisting in nuclear risk reduction, encourage discussions on ways in which emerging technologies can benefit disarmament, and underscore that, pending

28. Heather Williams, *Remaining Relevant: Why the NPT Must Address Emerging Technologies*, (London: Centre for Science and Security Studies, King's College, August 2020), https://www.kcl.ac.uk/csss/assets/remaining-relevant-new-technologies.pdf.
29. Williams, n. 16.
30. United Nations, *UN Web TV*, (17th Meeting) 3rd Preparatory Committee for Non-Proliferation of Nuclear Weapons Treaty Review Conference 2026, May 8, 2025, 35:25, http://webtv.un.org/en/asset/k1x/k1xki9tivc.

> the total elimination of nuclear weapons, nuclear-weapon States must reaffirm their commitment to maintain human control over decisions to use nuclear weapons and their delivery systems, while taking into account the need for the development of an international law regime around emerging technologies.[31]

AI is inherently political, and approaching any such governance through a rational and technical standpoint raises ethical questions about algorithmic biases, global power asymmetries resulting from AI uses, hegemonic forms of knowledge generation and use, among others.

The revised paragraph emerged following interventions by several countries that emphasised the importance of evaluating not only the risks, but also the potential benefits and opportunities offered by AI. El Salvador stated that the "duality" of these technologies needs to be better explored.[32] In such a situation, a potential starting point for dialogue is to consider the benefits from the technical applications of AI. As stated by El Salvador and echoed by many others, taking a less contentious, technical approach to the applications of AI in supporting not just the NPT, but also the broader non-proliferation regime, is a way to "depoliticise" AI.

Calls to depoliticise discussions on AI have also emerged in other multilateral settings. Some commentators, for instance, have described the REAIM initiative, mentioned earlier, as marking the first step towards a more inclusive and depoliticised approach to military AI governance. They state that "until the UN assumes a central role in this regard, addressing the challenges associated with regulating military AI becomes more attainable by aligning with the democratic, depoliticised, and decentralised principles

31. Reaching Critical Will, "Revised Draft Recommendations to the Review Conference of the NPT," NPT/CONF.2026/PC.III/CRP.4/Rev.1, May 8, 2025, https://reachingcriticalwill.org/images/documents/Disarmament-fora/npt/prepcom25/documents/CRP4Rev1.pdf.
32. n. 30.

articulated by REAIM."[33] While these claims are regarding the military applications of AI, they, nonetheless, reflect an interest in exploring the separation of technical AI cooperation from broader strategic rivalry.

Similar proposals have been made in the context of nuclear disarmament verification. Researchers like Dr Kim Westrich-Fellner have proposed that in the current security environment of increasing political polarisation that is compromising constructive efforts to advance nuclear disarmament, "states should explore more common ground in less politically contested areas. In this context, the current multilateral initiatives in Nuclear Disarmament Verification (NDV) offer a suitable platform for states to establish a baseline of cooperation. In particular, the technical, unpoliticised and collaborative set-up of NDV forums has the unique potential to generate consent on conceptual issues and objectives among a vast number of states, offering a path towards more effective policy-making in the NPT cycle."[34]

A recent May 2025 paper by the Oxford Martin AI Governance Initiative (AIGI) identifying the technical risks specific to international cooperation on AI safety research begins by stating that "international cooperation is common in AI research, including between geopolitical rivals."[35] Depoliticising AI has also been recommended in the context of AI safety research where "scientist-to-scientist collaboration has the potential to not only depoliticise global AI safety conversations, but also to improve the global inclusivity of these conversations."[36] Indeed, a technical and depoliticised approach to AI within the NPT is feasible and worth pursuing. Existing precedents in the field of

33. Mahmoud Javadi and Michal Onderco, "What Does Global Military AI Governance Need?" European Leadership Network, February 2, 2024, https://europeanleadershipnetwork.org/commentary/what-does-global-military-ai-governance-need/.
34. Kim Westerich-Fellner, "Nuclear Disarmament Verification and the NPT: De-Politicising the Political," European Leadership Network, January 9, 2024, https://europeanleadershipnetwork.org/commentary/nuclear-disarmament-verification-and-the-npt-de-politicising-the-political/.
35. Ben Bucknall, Saad Siddiqui, et al., *In Which Areas of Technical AI Safety Could Geopolitical Rivals Cooperate?*, (Oxford Martin School, AI Governance Initiative, May 2025), https://aigi.ox.ac.uk/wp-content/uploads/2025/05/In-Which-Areas-of-Technical-AI-Safety-Could-Geopolitical-2025-May-23.pdf.
36. Isabella Wilkinson, "Where Scientists Can Move the Needle on Global AI Safety Collaboration," *Revista de Prensa*, July 19, 2024, https://www.almendron.com/tribuna/where-scientists-can-move-the-needle-on-global-ai-safety-collaboration/.

AI (as identified in the AIGI paper) show that cooperation is possible despite geopolitical tensions and rivalry.

It can be argued that 'depoliticising AI' is an oxymoron. AI is inherently political, and approaching any such governance through a rational and technical standpoint raises ethical questions about algorithmic biases, global power asymmetries resulting from AI uses, hegemonic forms of knowledge generation and use, among others. Its development, governance, and deployment are driven by political and economic considerations such as state interests, private sector incentives, and geopolitical competition. For this paper, depoliticising AI does not imply that countries do not have political differences; rather, it provides an opportunity to explore common ground on technical aspects as the starting point for dialogue *despite* these political differences. AI *is* political, and "focusing on clearly defined, technically grounded topics – such as the use of AI in satellite imagery analysis for verification, or dispute mechanisms for issues that may arise from such applications"[37] does not erase such politics. Instead, such an approach to depoliticising AI could lay the groundwork for future engagement on the more politically sensitive aspects of AI, if the necessary political will emerges over time.

As suggested above, one practical starting point is to explore how AI can support nuclear verification, or safeguards (similar to Dr Kim Westrich-Fellner's suggestion on disarmament verification). This provides an example of exploring the potential benefits of AI, as suggested by several countries at the most recent NPT meeting, that can be meaningfully discussed in a depoliticised manner. Other suggestions for exploring the benefits of AI include its applications in nuclear risk reduction, disarmament, and peaceful uses of nuclear technology. For example, work in nuclear disarmament verification has already demonstrated how AI can enhance transparency, verifiability and irreversibility in disarmament efforts.[38] This paper suggests exploring the opportunities presented by AI in safeguards as a starting point

37. Rand, n. 27.
38. Stockholm Initiative for Nuclear Disarmament, "Navigating the Potential Impact," NPT/CONF.2026/PC.III/WP.35.

The potential of AI to improve and advance disarmament, arms control, risk reduction and the peaceful use of energy and technology should be actively considered as complementary avenues for depoliticising AI in the NPT.

for dialogue, since AI is already influencing how safeguards are implemented and presents a timely, technically-grounded entry point for engagement. This suggestion does not seek to prioritise one pillar of the NPT over the others, nor does it imply that the exploration of the impacts of AI in other areas should be excluded or take a back seat. Rather, the potential of AI to improve and advance disarmament, arms control, risk reduction and the peaceful use of energy and technology should be actively considered as complementary avenues for depoliticising AI in the NPT.

AI IN SAFEGUARDS

The International Atomic Energy Agency (IAEA) is an autonomous international organisation within the United Nations system[39], responsible for monitoring international civil nuclear power programmes and ensuring that nuclear materials from these programmes are not diverted from their peaceful purposes to make nuclear weapons. It does so through a set of technical measures to independently verify a country's legal commitment not to divert the said material to making nuclear weapons or other nuclear explosive devices.[40] While the IAEA is not a party to the NPT, it has been entrusted with specific responsibilities: under Article III of the NPT, the IAEA has a legal verification obligation as the international safeguards inspectorate. Countries accept these technical measures through the conclusion of safeguards agreements with the IAEA. IAEA safeguards are based on state declarations, on-site inspections by Agency personnel

39. International Atomic Energy Agency, "United Nations System," https://www.iaea.org/about/partnerships/united-nations-system. Accessed on May 23, 2025.
40. International Atomic Energy Agency, *IAEA Safeguards: Serving Nuclear Non-Proliferation*, June 2015, accessed April 13, 2025, https://www.iaea.org/sites/default/files/safeguards_web_june_2015_1.pdf.

and a variety of other information (open source and third party information). The technologies and methods used to collect and analyse this information are constantly evolving. In this context, the IAEA has recognised the potential of AI to provide important benefits to the implementation of safeguards, and has expressed interest in further using this technology.[41] In an era of a "new nuclear renaissance," which means an increased workload for inspecting new power programmes, and stagnant regular budgets, the IAEA "needs scalable inspection and safeguarding methods… Trying to find ways to do things better with automation and geospatial information is a hot topic of interest for the IAEA".[42]

It is well-understood that the volume of digital data received by the IAEA only increases over time. While such diversity is extremely beneficial and important for the IAEA in assessing whether a state is in compliance with its safeguards commitments, it also places a huge strain on analysts in collecting, processing and analysing such information.

It is well-understood that the volume of digital data received by the IAEA only increases over time. There is a large number and diversity of sources – including textual descriptions of declared research activities and site maps, satellite imagery, scanned reports and documents, a variety of audio and video recordings, images, scientific and technical publications, news reports, databases and social media that are available as unstructured data. While such diversity is extremely beneficial and important for the IAEA in assessing whether a state is in compliance with its safeguards commitments, it also places a huge strain on analysts in collecting, processing and analysing such

41. Cristina Siserman-Gray, Jonathan Barr, Jessica Burniske, Pegah Eftekhari, Robert Marek, and Aubrey Means, *Regulatory Challenges Related to the Use of Artificial Intelligence for IAEA Safeguards Verification*, Pacific Northwest National Laboratory (PNNL), https://resources.inmm.org/sites/default/files/2023-07/finalpaper_379_0512065638.pdf. Accessed on March 8, 2025.
42. Scott Stewart, Vandy Tombs, Jeremy Patterson, and Mike Channer, "Nuclear Safeguards, Security in Information Age Requires Integrated Approach," National Security Sciences Directorate, Oak Ridge National Laboratory, September 13, 2023, https://www.ornl.gov/news/nuclear-safeguards-security-information-age-requires-integrated-approach.

information. In such a scenario, AI can contribute significantly in various domains, particularly in analysing unstructured data received from open-source information.[43]

There are several ways in which AI can support safeguards implementation, though to some extent AI technologies are already being used. The Robotic Cherenkov Viewing Device (RCVD) is an example of this technology. It is an "automated surface vehicle used to verify spent nuclear fuel rods stored in spent fuel pools,"[44] though it has not yet achieved full autonomy. Advances in AI, particularly in natural language processing and transcription and translation, can ease the burden on analysts who often rely on keyword-based or search methods to sift through open source information. These methods can produce irrelevant results or miss critical data—simple queries often pull in irrelevant data (false positives), while complex ones need refinement and can still miss critical information (false negatives)—but advances in AI can help with processing audio, video and non-English sources.[45] Similarly, the analysis of satellite imagery is currently a highly manual task and limited resources generally only allow focussing on a subset of all facilities. This analysis can be enhanced by deep learning models that identify potential changes to infrastructure: "These sources of imagery associated with new techniques allow analysts to provide an in-depth assessment of nuclear-related facilities to support the state evaluation process and fulfil the IAEA's verification requirements more effectively."[46] Similarly, with over 1,300 surveillance cameras maintained by the IAEA at nuclear facilities around the world (as of 2021)[47], AI has the potential to dramatically reduce the time and effort needed to review footage, providing

43. P. Schneeweiss, T. Stojadinovic, and S. Baude, *The IAEA's Innovative Approach to Address the Challenges in the Collection and Analysis of Safeguards Relevant Information*, International Atomic Energy Agency, https://cdn.fourwaves.com/static/media/formdata/b12613f9-28b7-4568-a069-5c7cfeff3c8e/8d5ee6c8-9afc-4bf0-812c-2f57cf07b9d9.pdf. Accessed on March 8, 2025.
44. Jennifer Wagman and Teodor Nicula-Golovei, "The Evolution of Safeguards Technology," *IAEA Bulletin* 63, no. 3, October 2022, International Atomic Energy Agency, https://www.iaea.org/bulletin/the-evolution-of-safeguards-technology.
45. Schneeweiss, et. al., n. 43.
46. Wagman and Nicula-Golovei, n. 44.
47. Ibid.

Agency inspectors with timely support. While this is not an exhaustive list, these examples demonstrate how AI can increase the efficiency of, and confidence in, safeguards implementation.

The uses of any new and emerging technology are not without challenges, and it is important to be mindful of the challenges posed by AI in assisting with safeguards implementation. Recognising the challenges posed by the use of AI in safeguards helps to ensure its responsible and transparent integration. It also enables potential risks to be identified early on, allowing mitigation strategies to be developed before trust in the safeguards is compromised. Understanding the risks and challenges when reviewing the use of AI in safeguards is of direct concern to the IAEA and its member states. Some of the potential concerns are:

- How can the confidentiality of sensitive training data and AI model outputs be protected?
- What are the legal implications of using AI to contribute to safeguards conclusions, and how could this affect their credibility?
- How can transparency and non-discrimination in AI algorithms be demonstrated to maintain stakeholder trust?
- Who bears legal and ethical accountability in the event of errors or failures in AI-assisted safeguards processes?
- What measures are needed to address concerns relating to ethics, cyber security, bias, and potential misuse of AI tools in safeguards?
- How should Agency inspectors be trained to responsibly and effectively use AI technologies?
- How to ensure the appropriate decommissioning or retirement of AI systems in an ethical manner to protect the privacy of member states' data?

These are some concerns that are highly technical, requiring specialised expertise, and possibly outside the scope of the NPT review process. Detailed discussions of such a technical nature may open a can of worms in questioning a process and technology that the IAEA is still exploring. Moreover, many

AI is political, and taking a depoliticalised approach to it would require some creative boundary-setting in the scope of the discussions. In the NPT, this would mean identifying specific technical applications that can be separated from broader geopolitical disputes.

NNWS remain wary of supporting initiatives that could be seen as strengthening the non-proliferation pillar without commensurate progress on disarmament (such concerns would also apply to the discussions on the benefits of AI in supporting nuclear risk reduction, where the NNWS are generally averse to strengthening risk reduction measures that the NWS could use as a substitute to pursuing disarmament). This could reinforce perceptions of imbalance in the NPT's implementation of its three pillars and undermine trust in the regime's fairness. Nonetheless, it is important for the NPT community to be aware of the technical concerns even if they are not entirely discussed within the forum. A better and in-depth understanding of the challenges naturally would lead to better, more informed consultations and negotiations with one another.

As acknowledged previously, AI is political, and taking a depoliticalised approach to it would require some creative boundary-setting in the scope of the discussions. In the NPT, this would mean identifying specific technical applications that can be separated from broader geopolitical disputes for the time being, and exploring ways for these applications to encourage collaboration rather than clashes. In the context of safeguards, it is important to first recognise that AI is not intended to replace human inspectors or the critical role of human judgment in review processes. Rather, AI serves as a "smart aid." The use of AI allows inspectors to make better use of their time in the field, focussing more on analytical tasks, rather than compiling reports or other repetitive tasks—which, in turn, increases the effectiveness and efficiency of safeguards.[48] Increasing the overall *efficiency* of safeguards is an outcome that should align with the interests of all

48. Ibid.

parties. By extension, supporting the IAEA in its efforts to responsibly use AI to increase the efficiency of safeguards should also be in the interest of all parties.

A shared commitment from member states to support the IAEA could help to reinforce confidence in the development and use of these tools.

Given that all AI applications carry risks and limitations, a responsible approach would be to recognise these challenges while supporting the substantial efforts already being made by the IAEA to address them. A shared commitment from member states to support the IAEA could help to reinforce confidence in the development and use of these tools. Such a commitment would support the Agency's mandate and could signal a constructive, low-risk avenue for cooperation at a time when achieving a broader consensus within the NPT is difficult. It could also provide an opportunity to gain insight into the perspectives of the states-parties on AI in non-military applications.

LOOKING AHEAD: 2026 REVIEW CONFERENCE

The Review Conference for the eleventh review cycle of the NPT is scheduled to take place in New York from April 27 to May 22, 2026.[49] Both the intersessional period and the conference itself provide an ideal opportunity to consider how beneficial applications of AI can be leveraged to advance the treaty agenda. In addition to the comprehensive recommendations provided by the Stockholm Initiative in its working paper[50], the following recommendations can be considered in the lead up to the 2026 Review Conference:

- Consider establishing an independent expert group of technical experts from civil society and academia, maintaining gender, geographical and generational diversity, to conduct thorough and rigorous research on various aspects of emerging technologies and their risks and opportunities, to be made available to all countries for reference and

49. UNODA, "Decision on the Dates and Venue," NPT/CONF.2026/PC.III/DEC.1.
50. n. 38.

shared understanding. Support a regular and timely sharing of research outcomes as well as an interactive discussion on their merits and concerns, potentially hosted by organisations such as the United Nations Institute for Disarmament Research (UNIDIR).

- Promote the integration of forecasting and foresight methods to help states-parties proactively prepare for a range of possible scenarios, identify both risks and opportunities related to emerging technologies like AI, and support more informed, strategic decision-making within the NPT framework.

Thailand specifically argued for the integration of foresight methodologies at the third NPT Preparatory Committee, stating that "emerging risks must be addressed with foresight. The increasing convergence of advanced technologies—including AI, cyber, and space technologies—with nuclear command, control and delivery systems, has created new dimensions of complexity and unpredictability. This evolving landscape demands a coordinated, comprehensive strategy to mitigate potential unintended consequences and prevent catastrophic escalation."[51]

- The president-designate should take a conscious and active approach towards the issue of AI, and other emerging technologies more broadly. During the intersessional period and at the conference itself, the president-designate of the Review Conference should hold informal consultations with the states-parties, IAEA and other relevant technical and scientific experts, with the specific aim of improving understanding of the risks and opportunities associated with emerging technologies. In response to calls made by several delegations at the third Preparatory Committee, the president-designate should consider convening focussed sessions on the benefits of AI applications, with the aim of increasing collaboration and

51. United Nations Office for Disarmament Affairs, *Statement by Thailand*, General Debate, April 29, 2025, Third Session of the Preparatory Committee for the 2026 Review Conference of the Parties to the Treaty on the Non-Proliferation of Nuclear Weapons, https://docs-library.unoda.org/Treaty_on_the_Non-Proliferation_of_Nuclear_Weapons_-Preparatory_Committee_for_the_Eleventh_Review_Conference_(2025)/Thailand_-_GD_0.pdf.

identifying potential areas of consensus. In this regard, the IAEA should be invited to present an overview of the benefits and opportunities of AI in aiding its work.

- The states-parties should commit to supporting the IAEA in advancing the responsible and appropriate integration of AI tools in its activities related to the fields of nuclear safety, security and safeguards.
- Due to the technical nature of AI and other emerging technologies, regular consultations (both formal and informal) between diplomats and scientific experts should be actively encouraged to ensure informed and balanced discussions. To this end, the NPT review process could incorporate a specific mandate for the assessment of scientific and technological developments that may have an impact on the NPT.

CONCLUSION

With the 2026 NPT Review Conference approaching, it is time for discussions on AI to become more concrete and formal within the treaty framework. Although sensitive and military applications, such as AI in nuclear command and control, may remain politically contentious and provoke resistance, this should not prevent the NPT community from talking about them, and also from identifying and pursuing alternative practical entry points. Several countries have explicitly addressed the need to identify not just the risks, but also the opportunities presented by AI at the third Preparatory Committee. A depoliticised approach to AI, based on its technical benefits, is a potential way forward. Exploring AI applications in IAEA safeguards, for example, presents a timely and relevant starting point. AI is already shaping how safeguards are implemented, and this area aligns with shared interests in transparency and efficiency. However, there are legitimate concerns, ranging from ethical and legal questions to risks around bias, privacy, cyber security, misuse and accountability. It would indeed be beneficial to acknowledge and understand these challenges, support the IAEA's efforts in seeking to address these, and encourage more thoughtful and informed conversations,

even if not all of these issues are addressed or discussed directly within the NPT forum. By promoting responsible dialogue on the benefits of AI during the intersessional period and at the Review Conference, the states-parties can demonstrate that the NPT is responsive to technological change and lay the groundwork for future cooperation.