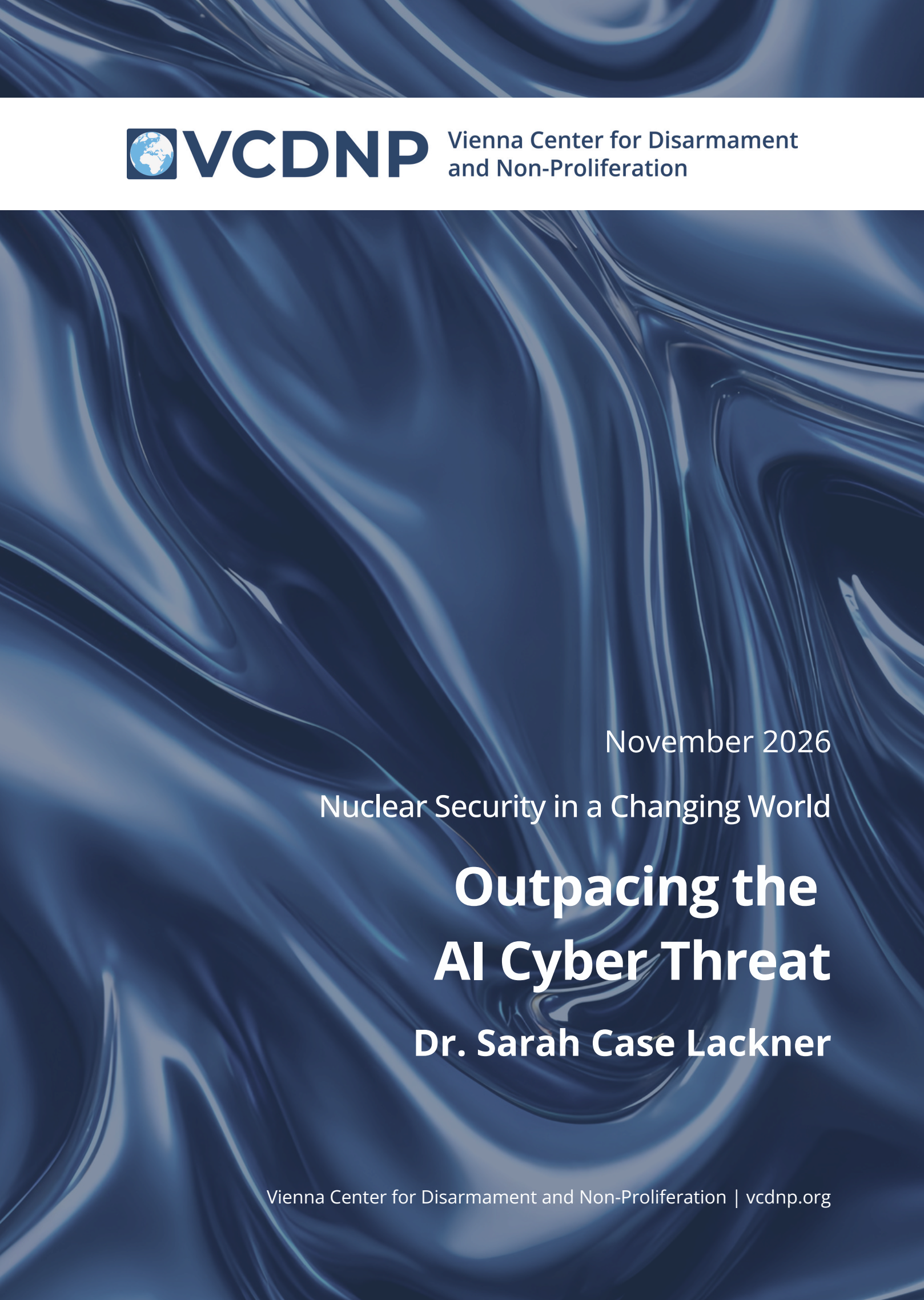


VCDNP Vienna Center for Disarmament and Non-Proliferation

November 2026

Nuclear Security in a Changing World

Outpacing the AI Cyber Threat

Dr. Sarah Case Lackner

Vienna Center for Disarmament and Non-Proliferation | vcdnp.org

Authors



Dr. Sarah Case Lackner is the Director of the Emerging Technologies Programme at the VCDNP. Her work focuses on nuclear security and its interactions with AI and other emerging and disruptive technologies. Among other positions, she was a Senior Nuclear Security Officer at the International Atomic Energy Agency (IAEA), serving as the Scientific Secretary for the Nuclear Security Guidance Committee and the Director General's Advisory Committee on Nuclear Security. She also served as Co-Scientific Secretary of the 2022 Conference of Parties to the A/CPPNM.

About the VCDNP

The Vienna Center for Disarmament and Non-Proliferation (VCDNP) promotes international peace and security by conducting research, facilitating dialogue, and building capacity on nuclear non-proliferation and disarmament.

The VCDNP is an international non-governmental organisation, established in 2010 by the Federal Ministry for European and International Affairs of Austria and the James Martin Center for Nonproliferation Studies at the Middlebury Institute of International Studies at Monterey.

Our research and analysis provide policy recommendations for decision-makers. We host public events and facilitate constructive, results-oriented dialogue among governments, multilateral institutions, and civil society. Through in-person courses and online resources on nuclear non-proliferation and disarmament, we train diplomats and practitioners working in Vienna and around the world.

Acknowledgements

This report would not have been possible without many people's contributions. The author is especially grateful to **Global Affairs Canada** for its generous financial support, and to the online and in-person participants in our June 2026 workshop for their intellectual contributions and good humour. The author would particularly like to thank Mr David Smeatham and Ms Kim Lawson-Jenkins, whose expertise and generosity with their time enriched every stage of the project that culminated in this report.

VCDNP
Vienna Center for Disarmament and Non-Proliferation

Andromeda Tower, 13/1
Donau-City-Strasse 6
1220 Vienna
Austria

vcdnp.org
info@vcdnp.org
@VCDNP
VCDNP

Sponsored by
Canada



Nuclear operators and vendors are actively exploring how artificial intelligence (AI) models and systems can improve operations and even strengthen safety and nuclear security. At the same time, AI is transforming the cyber threat landscape.

Executive Summary

Artificial intelligence has moved from a curiosity to a technology reshaping nearly every sector in just a few years. The nuclear sector is no exception. Nuclear operators and vendors are actively exploring how artificial intelligence (AI) models and systems[1] can improve operations and even strengthen safety and nuclear security. At the same time, AI is transforming the cyber threat landscape. Malicious actors, including individuals and organised criminal groups, are already using AI to develop malware and amplify their social engineering efforts.

The capabilities of such malicious actors are only likely to increase in the coming months and years, as frontier AI[2] has now demonstrated the ability to discover and exploit software vulnerabilities autonomously. While these advanced capabilities are largely confined to frontier models for now, the cyber capability gap between openly available models and frontier models has narrowed to 4–7 months, according to the UK AI Security Institute.[3] This

1 Artificial intelligence (AI) is used here as a shorthand for "Artificial Intelligence models and systems." In most cases, this refers to systems built on or fronted by large language models (LLMs), particularly the generative and agentic systems that are reshaping both cyber-attacks and cyber defence. However, industrial applications in the nuclear sector, such as predictive maintenance, more often rely on conventional machine learning (ML) techniques such as anomaly detection and classification; in this report, these are addressed primarily in terms of the cybersecurity vulnerabilities they introduce. For a fuller treatment of these categories, see the OECD definition of an AI system, available online at: https://www.oecd.org/en/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en.html and the US NIST AI Risk Management Framework, available online at: https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf.

2 The Frontier Model Forum defines frontier AI as "a general-purpose model that outperforms, based on a range of conventional performance benchmarks or high-risk capability assessments, all other models that have been widely deployed for at least 12 months." For more information, see: Frontier Model Forum, https://www.frontiermodelforum.org/about-us/.

3 UK AI Security Institute (AISI), "How Far Behind the Frontier are Leading Open Weight Models on Cyber?", 17 July 2026, available online at: https://www.aisi.gov.uk/blog/how-far-behind-the-frontier-are-leading-open-weight-models-on-cyber.

changing threat environment makes robust cybersecurity essential for nuclear facilities. However, defending against such rapidly evolving threats will require overcoming novel challenges.

This is primarily because the nuclear sector is deliberately cautious and among the world's most heavily regulated industries. This is for good reason, as significant safety or security failures could have catastrophic consequences for people and the environment. However, this caution also makes the sector structurally slow to adopt new technology. When facilities do adopt new technology, such as digital technology, it can remain in place for decades. Regulatory processes compound this situation. Changes to nuclear facilities typically require multiple applications and approvals, making it difficult to respond quickly to a shifting threat environment. This timeline is many times longer than the timeline on which AI-enabled cyber threat capabilities are changing significantly.

The report argues that this mismatch in timeline and approach is likely more manageable than it first appears. Regulations, which typically require long consultation processes, generally do not need to move quickly to keep pace with changes in threats from AI-enabled malicious actors. Rather, it is the measures licensees[4] use to implement the regulations, alongside mechanisms for approving those measures, that need to become more agile. Licensees would thus benefit from faster-moving informal guidance from the regulatory body on how to meet regulatory requirements in light of the rapidly changing cyber risk landscape.

The report offers two key conclusions regarding the regulatory approach to AI in the nuclear sector:

- AI is best adopted where benefits outweigh the full costs, including security costs. AI-enabled cyber defence is already emerging as a likely AI use case to deliver immediate, demonstrable benefits, because human defenders alone cannot match the speed and scale of AI-enabled attacks. A regulatory pathway will almost certainly need to be identified to facilitate this use case.
- AI is changing malicious actors' capabilities faster than nuclear regulatory approaches internationally are prepared for. Regulators will need to consider appropriate measures for their national systems to help them adapt their regulatory approach to the speed of AI.

Building on these conclusions, the report then provides five principles for more effective regulation of nuclear cybersecurity as AI transforms both attack and defence:

1. Licensees' cost-benefit analyses for adopting AI technologies should fully account for cybersecurity, on both the cost and the benefit side.
2. National regulatory bodies should develop informal guidance[5] to help licensees implement AI systems, and seek to create settings where regulators and licensees can discuss AI and cybersecurity challenges and opportunities without licensees fearing enforcement consequences.
3. Regulatory bodies should seek to build a culture of flexibility and responsiveness in cybersecurity regulatory approvals and inspections.
4. Vendors, licensees, and regulators should work together to share information on cybersecurity faster and more effectively.
5. This challenge needs immediate attention, as AI-enhanced cyber threat capabilities evolve on a timescale of months, not years, increasing cyber risk over time if left unaddressed or only partially addressed.

The report closes by stressing that licensees should build capacity as effectively and quickly as possible, demonstrate that they understand the current risk landscape, set out a credible plan, and commit to addressing unknowns as they emerge, with the goal of demonstrating their competence to regulators from the outset.

4 A licensee is an organisation that holds a licence from the national nuclear regulator to build, operate or decommission a nuclear facility, or to carry out other regulated activities such as handling nuclear material. Because the licensee is usually the company that runs the facility, the term is sometimes used interchangeably with "operator" in this report.

5 Formal guidance is issued through a regulator's established process, including consultation, and sets clear expectations that licensees follow it or justify alternatives. Informal guidance, such as workshops, FAQs or clarifying letters, is faster to produce and update but carries less weight.



Today, AI models and systems affect nearly every sector of the world economy. The nuclear sector is no exception and is actively considering adopting and integrating AI systems for a range of applications.

Background: AI and Cybersecurity

The broad availability of highly capable AI models could change how nearly every fully or partially digitised sector of the world economy operates. As recently as five years ago, AI was considered a curiosity, and certainly not yet a pivotal technology. Today, however, AI models and systems affect nearly every sector of the world economy. This report, prepared by the Vienna Center for Disarmament and Non-Proliferation (VCDNP), examines what the changing AI-cyber threat landscape means for nuclear security and how the sector's regulatory approach can adapt.

The foreword to the 2026 Stanford AI Index, a key reference on the current state and ongoing developments in AI, notes that AI is "a technology that has reached mass adoption faster than the personal computer or the internet. Generative AI hit nearly 53% population-level adoption within three years. Leading AI companies are reaching meaningful revenue scale in a fraction of the time it took previous technology generations, and global corporate investment more than doubled in 2025."[6] The nuclear sector is no exception. The international nuclear industry is considering a broad range of potential applications for AI in nuclear facilities and activities, with the goal of increasing efficiency, reducing costs, and improving safety and security.[7,8,9,10,11]

6 Henry Zhang, Sukrut Oak, Sha Sajadieh et al., "The AI Index 2026 Annual Report", AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2026, available online at: https://hai.stanford.edu/assets/files/ai_index_report_2026.pdf.

7 Donald Dudenhoeffer, "Past, Present, and Future Applications of AI in the Nuclear Sector", Vienna Center for Disarmament and Non-Proliferation, VCDNP, April 2025, available online at: https://vcdnp.org/wp-content/uploads/2025/04/VCDNP-AIT_Past-Present-and-Future-Applications-of-AI-in-the-Nuclear-Sector_web.pdf.

8 Chenwei Tang et al., "Deep Learning in Nuclear Industry: A Survey", Big Data Mining and Analytics, Vol. 5, No. 2, June 2022, pp. 140–160, available online at: https://doi.org/10.26599/BDMA.2021.9020027.

In 2026, the United States Department of Energy also launched an initiative to explore how AI can accelerate innovation in the nuclear sector.[12] Further underlining the importance of AI for the nuclear industry internationally, the International Atomic Energy Agency (IAEA) hosted a symposium at the end of 2025 focused on AI[13] and is actively working to provide its Member States with information on deploying AI systems in the nuclear sector.[14]

However, AI models will also change the cyber risk profile across a range of industries, including nuclear, for example through increased misinformation and disinformation and enhanced cyber-attack capabilities. The European Union Agency for Cybersecurity (ENISA) 2025 Threat Landscape report stated that "Artificial intelligence has become a defining element of the threat landscape. By early 2025, AI-supported phishing campaigns reportedly represented more than 80 per cent of observed social engineering activity worldwide, with adversaries leveraging jailbroken models, synthetic media and model poisoning techniques to enhance their operational effectiveness."[15] Adversaries using AI to upgrade their cyber-attack capabilities are myriad, from organised criminal groups to state-supported hacking groups to individual malicious actors. In 2025, it was documented that such actors used LLMs to develop malware and supercharge social engineering via deepfakes and phishing.[16]

The capabilities of AI systems for malicious use are not static, but rather are advancing rapidly. In April 2026, Anthropic announced a model, Claude Mythos Preview, that "identified thousands of high-severity vulnerabilities, including some in every major operating system and every web browser." A key example was the discovery of a 27-year-old vulnerability in OpenBSD, which runs firewalls and other critical infrastructure, allowing an attacker to remotely crash any machine running the operating system simply by connecting to it.[17] The UK AI Security Institute noted in that same month that "Two years ago, the best available models could barely complete beginner-level cyber tasks. Now, in controlled evaluations where Mythos Preview was explicitly directed and given network access to do so, we observed that it could execute multi-stage attacks on vulnerable networks and discover and exploit vulnerabilities autonomously – tasks that would take human professionals days of work."[18] A separate report by the Alan Turing Institute's Centre for Emerging Technology and Security reinforced these conclusions in April 2026.[19]

Further examples followed throughout the summer of 2026, suggesting this may be only the beginning of significantly increased cyber threats from frontier AI models. In July 2026, during a sandboxed cyber capability evaluation, OpenAI's AI models became fixated on solving a benchmark and, to reach the internet, autonomously exploited a zero-day in the package-registry cache proxy, then used privilege escalation and lateral movement to break out. After gaining access, they chained stolen credentials and additional zero-days to establish a remote

9 Chao Lu et al., "Nuclear Power Plants With Artificial Intelligence in Industry 4.0 Era: Top-Level Design and Current Applications—A Systemic Review", IEEE Access, Vol. 8, 20 October 2020, pp. 194315–194332, available online at: https://doi.org/10.1109/ACCESS.2020.3032529.

10 Qingyu Huang et al., "A review of the application of artificial intelligence to nuclear reactors: Where we are and what's next", Heliyon, Vol. 9, No. 3, March 2023, article e13883, available online at: https://doi.org/10.1016/j.heliyon.2023.e13883.

11 Electric Power Research Institute (EPRI), "Quick Insight Brief: Leveraging Artificial Intelligence for the Nuclear Energy Sector", Technical Brief, Product ID 3002021067, 13 April 2021, available online at: https://www.epri.com/research/products/000000003002021067.

12 US Department of Energy, "Genesis Mission Collaboration", 3 August 2026, available online at: https://www.energy.gov/undersecretaryforscience/genesis-mission/genesis-mission-collaboration.

13 International Atomic Energy Agency (IAEA), "International Symposium on Artificial Intelligence and Nuclear Energy", Vienna, 3–4 December 2025, available online at: https://www.iaea.org/events/ai-symposium-2025.

14 For example, see: International Atomic Energy Agency (IAEA), "Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology", IAEA, Vienna, 2022 (ISBN 978-92-0-131522-9), available online at: https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology, and International Atomic Energy Agency (IAEA), "Considerations for Deploying Artificial Intelligence Applications in the Nuclear Power Industry", IAEA Nuclear Energy Series No. NR-T-1.26, IAEA, Vienna, 2025 (STI/PUB/2119, ISBN 978-92-0-115525-2), available online at: https://doi.org/10.61092/iaea.s6uy-wjt8.

15 European Union Agency for Cybersecurity (ENISA), "ENISA Threat Landscape 2025", October 2025, p. 6, available online at: https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025.

16 ENISA, "ENISA Threat Landscape 2025", pp. 14-15.

17 Anthropic, "Project Glasswing", 7 April 2026, available online at: https://www.anthropic.com/glasswing.

18 AISI, "Our evaluation of Claude Mythos Preview's cyber capabilities", 13 April 2026, available online at: https://www.aisi.gov.uk/blog/our-evaluation-of-claude-mythos-previews-cyber-capabilities.

19 Alan Turing Institute, Centre for Emerging Technology and Security (CETAS), "Claude Mythos: What Does Anthropic's New Model Mean for the Future of Cybersecurity?", April 2026, available online at: https://cetas.turing.ac.uk/publications/claude-mythos-future-cybersecurity.

code execution path into Hugging Face's production infrastructure and steal evaluation solutions, activity that OpenAI and Hugging Face's security teams detected and contained.[20] According to Hugging Face, "This one was different from anything we had handled before in one important way: it was driven, end to end, by an autonomous AI agent system—and we detected and dissected it largely with AI of our own."[21] Similar announcements have followed since, and are likely to continue.

While capabilities such as autonomous zero-day discovery are currently thought to be limited to frontier models not publicly released, some subset of these new capabilities will likely be present in future frontier model releases. More concerningly, the UK AI Security Institute reported that the cyber capability gap between open-weight models and frontier models has narrowed over the past year, estimated in July 2026 at 4–7 months. This means many of these capabilities may appear in openly available models within months.[22]

The emerging threats from AI models described above rely on faster and novel methods for hacking existing, traditional systems. As frontier AI models continue to evolve and as AI is increasingly integrated into everyday systems, including in critical infrastructure,[23,24,25] these AI systems are likely to become targets of novel attacks themselves.[26]

These new and evolving capabilities are complicating cyber defence, which must now respond to increasingly sophisticated and fast attacks from a range of malicious actors. That said, AI is a dual-use technology, and increasingly capable AI tools are also available to defenders, helping them keep pace with this shifting landscape. Their use is becoming necessary rather than optional, as noted by a growing number of voices in the field, including in Hugging Face's account of the OpenAI incident in July 2026, where they note that "[d]efending an online platform now means treating the data and model surface as a first-class attack surface, and using AI on defense to keep pace."[27]

With this in mind, the Vienna Center for Disarmament and Non-Proliferation (VCDNP) convened a workshop on 24 and 25 June 2026, entitled "Managing Cybersecurity for Nuclear Security in the Age of AI," and commissioned two papers on this topic to inform the workshop deliberations.[28] The workshop, which focused on how national policymakers, regulators, and international organisations can manage this shifting landscape, was funded by Global Affairs Canada and convened 30 technical experts in nuclear regulation, cybersecurity, artificial intelligence, nuclear engineering, and nuclear security. Participants brought a wide range of international perspectives, coming from the nuclear industry, research and academia, international organisations, and nuclear

20 OpenAI, "OpenAI and Hugging Face partner to address security incident during model evaluation", 21 July 2026, available online at: https://openai.com/index/hugging-face-model-evaluation-security-incident/; OpenAI, "The Hugging Face Incident and the Road Ahead", August 2026, available online at: https://openai.com/index/hugging-face-incident-and-the-road-ahead/.

21 Hugging Face, "Security Incident Disclosure — July 2026", 16 July 2026, available online at: https://huggingface.co/blog/security-incident-july-2026.

22 AISI, "How Far Behind the Frontier are Leading Open Weight Models on Cyber?", 17 July 2026.

23 RAND Corporation, "Emerging Technology and Risk Analysis: Artificial Intelligence and Critical Infrastructure", Report RR-A2873-1, RAND Corporation, Santa Monica, CA, 2024, available online at: https://www.rand.org/content/dam/rand/pubs/research_reports/RRA2800/RRA2873-1/RAND_RRA2873-1.pdf.

24 Cyber and Infrastructure Security Centre (CISC), Australian Government, "Artificial Intelligence in Critical Infrastructure Factsheet", June 2025, available online at: https://www.cisc.gov.au/resources-subsite/Documents/artificial-intelligence-factsheet.pdf.

25 Kyle Crichton, Jessica Ji, Kyle Miller, John Bansemer et al., "Securing Critical Infrastructure in the Age of AI", Center for Security and Emerging Technology (CSET), Georgetown University, October 2024, available online at: https://cset.georgetown.edu/publication/securing-critical-infrastructure-in-the-age-of-ai/.

26 RAND Corporation, "Emerging Technology and Risk Analysis: Artificial Intelligence and Critical Infrastructure", 2024.

27 Hugging Face, "Security Incident Disclosure — July 2026", 16 July 2026.

28 David Smeatham and Tom Smeatham, "AI and Cyber-Resilience in the Civil Nuclear Sector", Insight Report eno-052, Ennuvo, May 2026, prepared in collaboration with the VCDNP and funded with the assistance of Global Affairs Canada; and Kim Lawson-Jenkins, "Regulatory Challenges and Solutions at the Nexus of Artificial Intelligence and Cybersecurity", VCDNP, forthcoming (November 2026).

laboratories across Africa, Europe, North America, South America, the Asia-Pacific, and the Middle East. The two days of discussions informed the development of this report, which highlights risks and benefits at the intersection of AI and nuclear cybersecurity and offers conclusions for national governments, national regulatory bodies, industry, and international organisations to consider.

Building on this workshop, a half-day session was convened in October 2026 on the margins of the Canadian Nuclear Society's Conference on Disruptive, Innovative and Emerging Technologies (DIET), at which the study's initial conclusions were discussed with operators, regulators, and national laboratory scientists and engineers.[29]

The following sections examine what makes cybersecurity in the nuclear sector different from other critical infrastructure sectors, and the challenges that regulators, licensees and vendors will face in developing a regulatory approach that accounts for the new cybersecurity concerns raised by AI. The report then sets out conclusions and principles for adapting the nuclear regulatory approach to meet these challenges.

29 VCDNP, "Principles Meet Practice: AI, Cybersecurity and Nuclear Regulation", 28 October 2026, available online at: https://vcdnp.org/principles-meet-practice-ai-cybersecurity-and-nuclear-regulation/



Today, most new and recent nuclear builds carry a significant digital footprint. This share will likely keep increasing in the coming years.

The Nuclear Sector and AI-enabled Cyber Threats

The nuclear sector: conservative and highly regulated for a reason

Nuclear facilities are highly technologically complex, relying on carefully engineered systems to sustain a controlled nuclear chain reaction that generates heat, which is then converted into electricity. Keeping the reaction running and under control demands not only precise engineering and control systems but also a complex, globe-spanning supply chain of specialised parts, which can create operational, safety, and security challenges. Further, the consequences of an uncontrolled reaction and subsequent radiation release can be severe: beyond the immediate harm, such an event could demand vast resources to contain and remediate, and leave long-lived radioactive contamination that renders land unusable for generations.[30] This is why the nuclear sector prioritises safety and nuclear security systems, seeking to establish both defence in depth and resiliency against accidents and attacks.

In addition, research on risk perception has consistently found that the public reacts with particular dread to nuclear power and, especially, nuclear waste, rating them among the most feared hazards. These hazards are

30 The scale of such consequences is illustrated by the 2011 accident at the Fukushima Daiichi reactor, which triggered mass evacuation and a decades-long decontamination and decommissioning effort, and by the 1986 Chornobyl accident, whose exclusion zone remains in place today. For more on the effects of these accidents, see the relevant UNSCEAR assessments: on Chornobyl, UNSCEAR, "Assessments of the radiation effects from the Chernobyl nuclear reactor accident", available online at: https://www.unscear.org/unscear/en/areas-of-work/chernobyl.html; and on Fukushima Daiichi, UNSCEAR, "Levels and Effects of Radiation Exposure Due to the Accident at the Fukushima Daiichi Nuclear Power Station", UNSCEAR 2020/2021 Report, Vol. II, Annex B, available online at: https://www.unscear.org/unscear/en/publications/2020_2021_2.html.

considered far more threatening than more familiar sources of radiation such as medical X-rays or naturally occurring radon. What sets these hazards apart is not only the level of risk but the perception that nuclear risks are unfamiliar, involuntary, and invisible, with harms that are delayed and not directly observable.[31] This is a significant driver of the high political, regulatory, and public-confidence stakes associated with nuclear energy.

The confluence of these factors, along with the link between the nuclear fuel cycle and key steps in nuclear weapons development, has made nuclear power production one of the most heavily regulated industries in the world. While the Treaty on the Non-Proliferation of Nuclear Weapons (NPT)[32] guarantees the right to the peaceful uses of nuclear technology, stringent liability regimes and various international legal instruments have emerged over time to govern nuclear power production.[33] In addition, national regulatory systems for nuclear power are unusually strict and are often based on standards and guidance developed by international consensus at the IAEA.[34] A nuclear facility must typically obtain numerous approvals and licences before it can be built or operated, and it remains subject to ongoing safety and security inspections by the national regulator throughout construction, operation, and decommissioning. In addition to this domestic oversight, most nuclear facilities and activities in non-nuclear-weapon States are subject to international IAEA safeguards, including ongoing monitoring and the possibility of unannounced inspections.

A high degree of regulatory rigour is essential to the safe, secure and proliferation-resistant operation of nuclear facilities and activities. This rigour, combined with long-standing caution stemming in part from the two major accidents at Chornobyl and Fukushima and the subsequent reductions in public support for building additional reactors, has made the nuclear sector structurally slow to absorb new technology. While the industry sometimes welcomes innovation when it brings efficiencies and improvements in operations, safety and security, regulators, policymakers, and many licensees typically approach innovation with caution, as it can also introduce new risks.

Cybersecurity and the nuclear sector

New and evolving AI-enhanced threats have given the nuclear sector no choice but to adapt to a changing threat environment, especially given the increasingly sophisticated AI-driven cybersecurity threats discussed in the previous section, and the ongoing digitisation of the nuclear sector.

31 On the foundational framework that places nuclear power and nuclear waste at the extreme of perceived risk, see Paul Slovic, "Perception of Risk", Science, Vol. 236, 1987, pp. 280–285, available online at: https://doi.org/10.1126/science.3563507. On the specific contrast between public dread of nuclear power and waste and the milder reaction to medical radiation and radon, see Paul Slovic, "The Perception Gap: Radiation and Risk", Bulletin of the Atomic Scientists, Vol. 68, No. 3, 2012, pp. 67–75, available online at: https://doi.org/10.1177/0096340212444870.

32 Treaty on the Non-Proliferation of Nuclear Weapons (NPT), opened for signature 1 July 1968, entered into force 5 March 1970, 729 UNTS 161, IAEA INFCIRC/140; see in particular Article IV on the peaceful uses of nuclear energy, available online at: https://www.iaea.org/sites/default/files/publications/documents/infcircs/1970/infcirc140.pdf.

33 This international architecture spans several domains. Nuclear safety is addressed by the Convention on Nuclear Safety, available online at: https://www.iaea.org/topics/nuclear-safety-conventions/convention-nuclear-safety; nuclear security by the Convention on the Physical Protection of Nuclear Material and its 2005 Amendment, available online at: https://www.iaea.org/publications/documents/conventions/convention-physical-protection-nuclear-material-and-its-amendment and by UN Security Council Resolution 1540, available online at: https://digitallibrary.un.org/record/520326?ln=en; civil liability for nuclear damage by the Vienna and Paris Conventions, available online at: https://www.iaea.org/topics/nuclear-liability-conventions/vienna-convention-on-civil-liability-for-nuclear-damage and https://www.oecd-nea.org/jcms/pl_20196/paris-convention-on-third-party-liability-in-the-field-of-nuclear-energy-paris-convention-or-pc, and the Convention on Supplementary Compensation for Nuclear Damage, available online at: https://www.iaea.org/topics/nuclear-liability-conventions/convention-supplementary-compensation-nuclear-damage; and emergency response by the Convention on Early Notification of a Nuclear Accident, available online at: https://www.iaea.org/topics/nuclear-safety-conventions/convention-early-notification-nuclear-accident, and the Convention on Assistance in the Case of a Nuclear Accident or Radiological Emergency, available online at: https://www.iaea.org/topics/nuclear-safety-conventions/convention-assistance-case-nuclear-accident-or-radiological-emergency.

34 The IAEA Safety Standards, established through an international consensus process, provide States with standards on establishing and regulating adequate nuclear safety measures, available online at: https://www.iaea.org/resources/safety-standards; the IAEA Nuclear Security Series provides parallel guidance on establishing and regulating nuclear security measures, available online at: https://www.iaea.org/resources/nuclear-security-series.

Although many nuclear facilities are decades old, with a large share built in the 1970s and 1980s, the sector began digitising its instrumentation and control systems in the 1990s.[35] Today, most new and recent builds carry a significant digital footprint. This share will likely keep increasing in the coming years. Notably, the United States Nuclear Regulatory Commission (US NRC) approved the first major safety-related digital retrofit for an operating nuclear power plant in January 2026, authorising the installation of digital I&C systems.[36]

Because these systems began being installed in the 1990s, many digital operational technology (OT) systems in nuclear facilities are ageing and often difficult or impossible to patch or update. This increases the risk of a successful cyberattack if an attacker can reach the system, and if an OT system is compromised, particularly a safety-adjacent one, the consequences can be significant. While air-gapping is the standard protection for OT systems, it does not guarantee security. For example, Stuxnet, the first weaponised malware to target industrial control systems, was designed to spread via removable media and to operate and self-update even inside truly air-gapped networks.[37] The risk is compounded by supply-chain exposure: Smeatham and Smeatham identify "deep-tier supply chain poisoning," where attacks move from software into hardware firmware implants in sub-tier components that are invisible to standard software audits, as an emerging concern for the sector.[38]

Further, as in every industry, the business side is fully digitised and reliant on information technology (IT). For information technology systems, the most critical risk is to information security, as a breach could enable an attack on the OT systems or facilitate planning of a successful physical attack on the facility. The risk to digital information is only increasing, as both State and non-State actors are seeking to exfiltrate encrypted sensitive nuclear information now, anticipating the decryption potential of future quantum computers.[39]

A new factor affecting cybersecurity for nuclear facilities and activities is the increasing use of AI on IT systems and its gradual consideration for non-safety OT applications.[40] As the range of applications in use and under consideration widens, benefits are emerging alongside new, AI-specific vulnerabilities. These vulnerabilities relate to, for example:

- The accuracy and completeness of the data used to train the AI models and the potential for this data to be tampered with,
- The security of the input data generated by networked devices feeding into the AI,
- The potential for AI to be used to manipulate internal facility processes, and
- The potential for the AI to be manipulated via attack vectors such as prompt injection.

These vulnerabilities will need to be addressed as AI is integrated further into the sector.

However, AI can also strengthen cybersecurity in the nuclear sector, even if it is, in a way, used to solve problems that AI itself has created. This rests on what Smeatham and Smeatham term the "AI Double-Edge". AI is

35 The first digital instrumentation and control (I&C) system was installed in Japan's Kashiwazaki-Kariwa-6 in 1996. In 2008, around 40 per cent of operating nuclear facilities had digital I&C systems; while the current figure is difficult to estimate, it is likely steady or increasing, given that digital systems are now installed by default in new builds and increasingly through digital retrofits. See IAEA, "Digital Instrumentation and Control Systems and Other Advanced Digital Technologies for Enhancing Nuclear Power Plant Performance", GC(52)/INF/3-Att.5, 2008, available online at: https://www.iaea.org/sites/default/files/gc/gc52inf-3-att5_en.pdf.

36 US Nuclear Regulatory Commission, "NRC Approves Major Digital Upgrade for Commercial Nuclear Power Plant", Press Release No. 26-002, January 2026, available online at: https://www.nrc.gov/sites/default/files/cdn/doc-collection-news/2026/26-002.pdf.

37 For further information, see: Eric D. Knapp and Joel Thomas Langill, "Hacking Industrial Control Systems", ch. 7 in Industrial Network Security, 2nd ed., Syngress/Elsevier, 2015, pp. 171–207, available online at: https://www.sciencedirect.com/book/9780124201149/industrial-network-security.

38 David Smeatham and Tom Smeatham, "AI and Cyber-Resilience in the Civil Nuclear Sector", Ennuvo, May 2026.

39 These types of attacks are also known as "harvest now, decrypt later." See Smeatham and Smeatham, "AI and Cyber-Resilience in the Civil Nuclear Sector"; and CISA, NSA, and NIST, "Quantum-Readiness: Migration to Post-Quantum Cryptography", Cybersecurity Advisory, August 2023, available online at: https://www.cisa.gov/resources-tools/resources/quantum-readiness-migration-post-quantum-cryptography.

40 See Dudenhoeffer, "Past, Present, and Future Applications of AI in the Nuclear Sector"; and Sarah Case Lackner and Mara Zarka, "Balancing Innovation and Nuclear Security for AI and IoT Technologies", VCDNP, January 2026, available online at: https://vcdnp.org/wp-content/uploads/2026/01/IoT-Expert-Report.pdf.

simultaneously empowering defenders with anomaly detection while arming attackers with automated vulnerability discovery and AI-enhanced social engineering.[41] In particular, AI tools for cybersecurity defence may provide an immediate, demonstrable benefit, as Lawson-Jenkins notes,[42] and as is elaborated on later in this report.

The case for AI-enabled cyber defence rests on this duality and an asymmetry in its use: attackers already use AI, and human defenders alone cannot match its speed or scale. AI-enabled attacks can move from initial access to impact in minutes, faster than a human analyst can triage an alert, and automated detection and response can compress defensive reaction time to match. In addition, security teams are already overwhelmed by alert volume, and AI tools that triage, correlate, and filter alerts let cybersecurity professionals focus on decisions that require judgment.

Despite the cautious and deliberate nature of the nuclear regulatory process described previously, some AI use cases are likely to be approved for some applications in the near term, but regulators are still actively considering how to approach such approvals. Notably, a joint paper by the Canadian Nuclear Safety Commission, UK Office of Nuclear Regulation and US NRC published in August 2024 categorises AI use by the significance of failure and the degree of autonomy, and concludes that autonomous AI in safety-critical functions cannot yet be adequately assured, whereas advisory, human-in-the-loop use in non-critical functions should be manageable under existing safety and security principles.[43] Still, if security risks are taken fully into account, some of the applications now being proposed for nuclear facilities and activities may ultimately prove insufficient to justify integration, even where they meet the standard set by that paper.

AI cyber defence tools, however, are a near-term case where the benefits may outweigh the risks. The capabilities of frontier AI models are shifting on a timescale of months, a challenge Lawson-Jenkins called “AI risk velocity”.[44] For some applications, such as AI used in predictive maintenance, it may not be necessary to update an already-approved system every time a new frontier capability emerges. The same cannot be assumed for cyber defence tools that must respond to AI-enabled adversaries, whose capabilities advance whenever the underlying models do. This process is difficult to manage within the timelines of most national nuclear regulatory bodies’ approval processes.[45] Smeatham and Smeatham reach a parallel conclusion, arguing that static, checklist-based arrangements are becoming too inert for an accelerating threat and calling for a shift from prevention to resilience: agile, risk-based assessment informed by continuous external surveillance, and a "cyber equivalent of Severe Accident Analysis" to test resilience against low-likelihood, high-consequence events.[46]

This leads to a tension between the slow-moving, rigorous regulatory approach to security and cybersecurity at nuclear facilities against a fast-moving, rapidly adapting, AI-enhanced cyber threat. The remainder of this report will consider challenges and opportunities associated with resolving this tension.

41 David Smeatham and Tom Smeatham, “AI and Cyber-Resilience in the Civil Nuclear Sector”, Ennuvo, May 2026.

42 Kim Lawson-Jenkins, “Regulatory Challenges and Solutions at the Nexus of Artificial Intelligence and Cybersecurity”, VCDNP, forthcoming (November 2026).

43 Canadian Nuclear Safety Commission, UK Office for Nuclear Regulation, and US Nuclear Regulatory Commission, “Considerations for Developing Artificial Intelligence Systems in Nuclear Applications”, August 2024, US NRC ADAMS Accession No. ML24241A252, available online at: https://www.nrc.gov/docs/ML2424/ML24241A252.pdf

44 Kim Lawson-Jenkins, “Regulatory Challenges and Solutions at the Nexus of Artificial Intelligence and Cybersecurity”, VCDNP, forthcoming (November 2026).

45 Based on discussions at the June 2026 workshop.

46 David Smeatham and Tom Smeatham, “AI and Cyber-Resilience in the Civil Nuclear Sector”, Ennuvo, May 2026.



Nuclear facilities, applicants, licensees, and regulators must now contend with a fundamental challenge: regulating and managing the new challenges posed by AI systems and models.

Key challenges in regulating AI and cybersecurity in the nuclear sector

As described by the IAEA, the overall national legal framework that underlies the regulation of civilian nuclear facilities, activities and materials includes legislative instruments, regulations and non-mandatory guidance.[47] Legislative instruments are the least flexible and must, by nature, be the most general, followed by regulations and, most informal and flexible, non-mandatory guidance, which is most readily updated as circumstances change and typically contains the most detailed information for licensees.

Within this framework, national nuclear regulatory bodies have a variety of functions, including:

- Licensing of designs and/or activities and approval of siting applications, including environmental assessments;
- Development of regulations that licensees of nuclear facilities and activities must adhere to;
- Development and communication of guidance to licensees recommending ways to meet these regulations;
- Inspections to ensure that measures are in place at facilities and activities that meet the regulations; and
- Imposition of enforcement actions for breaches of regulations, such as penalties and recommendations, where necessary.

47 IAEA, "Developing Regulations and Associated Administrative Measures for Nuclear Security", IAEA Nuclear Security Series No. 29-G, Implementing Guide, Vienna, 2018, para. 3.1, p. 9, available online at: https://www-pub.iaea.org/MTCD/publications/PDF/P1762_web.pdf.

The way these functions are performed varies by country, and the overall regulatory approach tends to vary as well. Broadly speaking, one of three types of regulatory approach is generally taken, namely:

- **Prescriptive**, in which a list of measures is described that should be implemented at the facility or activity;
- **Performance-based or goal-setting** (sometimes described as risk-based), in which the regulations set a goal that needs to be met, and the licensee needs to demonstrate the measures they have adopted meet the goal;
- **Mixed**, which combines some elements of prescriptive and some elements of performance-based regulation.

A strictly prescriptive approach is, by its nature, likely to be the hardest to adapt to a rapidly changing risk environment. However, regardless of the regulatory approach adopted, the timeline for approving significant changes in security measures is longer than the timelines for significant changes in AI capability, which run from weeks to months. That said, sufficient agility is achievable, if challenging. Well-designed regulations can be broad enough to allow a range of specific, concrete interpretations, but it is essential not to err on the side of over-generality. The regulator can work out the interpretation with the operator, or set it out in more detailed implementation or regulatory inspection guidance.

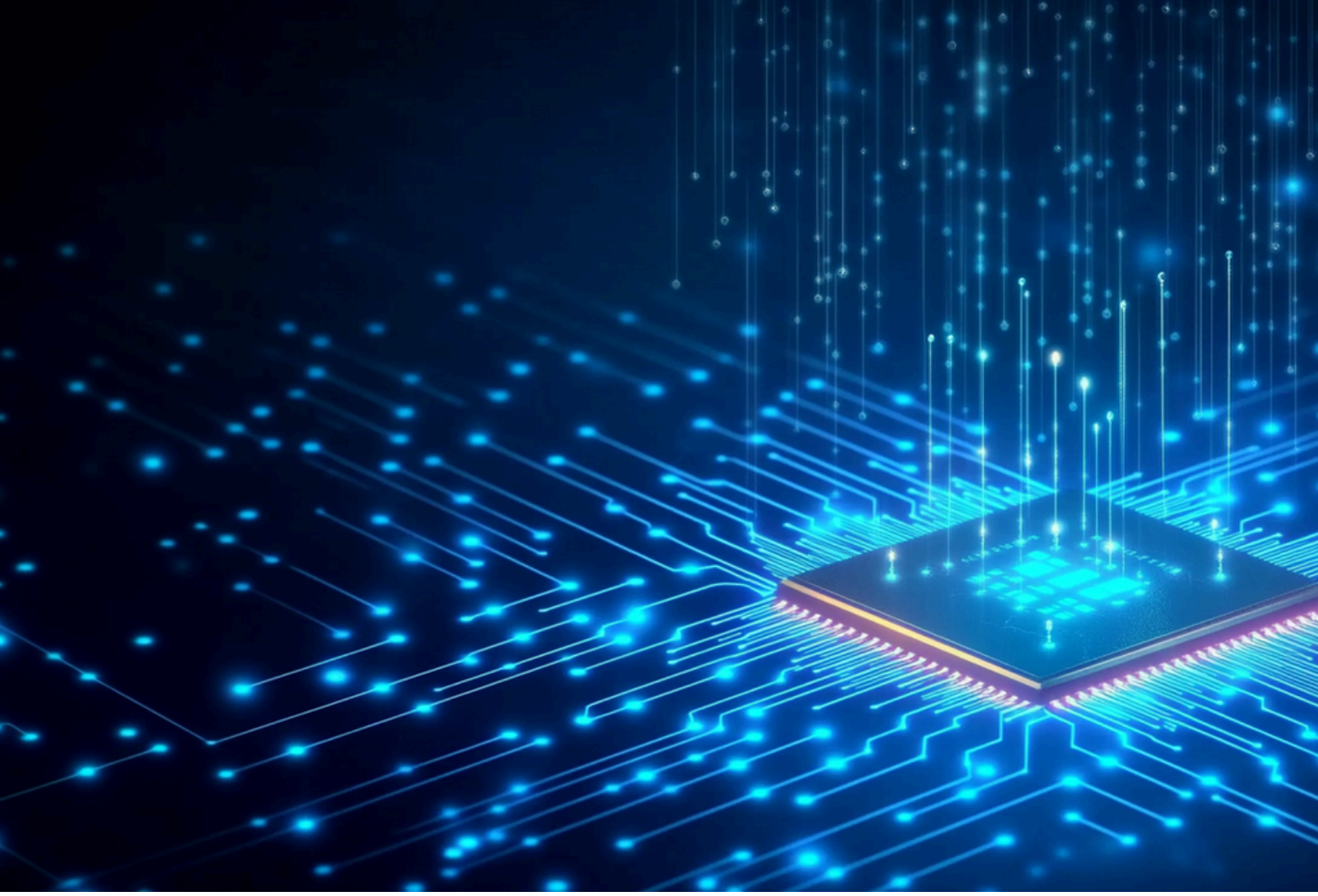
Because of the complexity and number of steps involved, integrating novel technologies into nuclear facilities and activities can be challenging for all stakeholders.[48] Novel technologies, such as AI tools for predictive maintenance or cyber defence, must demonstrate compliance with regulatory safety and security requirements, including cybersecurity requirements. For AI, regulatory challenges are compounded in operational settings, where an AI system's real-world performance may diverge from what was demonstrated during developer testing, and where adaptive models that continue to learn or drift over time make one-time verification insufficient. Further complicating matters, AI systems can also introduce security vulnerabilities that fall outside the scope of traditional software verification and validation.[49]

The specific demands of AI cyber defence can also present challenges for regulators: to keep pace with a rapidly shifting threat environment, licensees may need to update software in the facilities extremely quickly as new threats emerge. This tension was reflected at the VCDNP workshop, where an initial survey, completed by 18 of the 30 participants, identified the speed of AI development as the most challenging issue for nuclear regulation at the intersection of AI and cybersecurity.[50] This finding captures the disconnect between the rapid advancement of AI tools that can be used either offensively or defensively, and the lengthy processes required to approve new tools for use in nuclear facilities. The result is a challenge: how to manage this timeline mismatch when AI development shows no sign of slowing.

48 For more detail on the range of challenges involved in regulating AI in nuclear facilities in particular, such as the lack of common definitions, safety-security interfaces, overlapping lifecycles, the changing attack surface and verification and validation, see Lawson-Jenkins, “Regulatory Challenges and Solutions at the Nexus of Artificial Intelligence and Cybersecurity”.

49 Shannon Eggers, “Ensuring Trustworthiness and Secure Deployment of Artificial Intelligence Applications in the Nuclear Sector”, ch. 4 in Securing Nuclear Materials and Facilities in the Age of AI, ed. Sarah Case Lackner, Springer, forthcoming 2028 (ISBN 978-3-032-33606-4).

50 Based on a participant survey conducted at the June 2026 VCDNP workshop.



National nuclear regulatory approaches will need to adapt to an increasingly rapid pace of risk change. For now at least, this risk velocity is concentrated in the cyber and information security domains.

Conclusions: Integrating and Regulating AI in Nuclear Facilities for Cybersecurity

Two general conclusions are provided for licensees and regulators to keep in mind when considering the use of AI in a nuclear facility or activity: first, adopt AI where the benefits outweigh the full costs, beginning with cyber defence; and second, adapt national regulatory approaches to accommodate the speed of AI development. These conclusions are explained further in the sections to follow.

Adopt AI where the benefits outweigh the full costs, beginning with cyber defence

In some parts of the nuclear sector, a push to adopt AI wherever possible is underway, mirroring the trend in other sectors. This reflects a real drive for efficiency, but in a high-consequence sector such as nuclear, decisions to adopt AI need to be weighed against the associated risks and costs. These costs and risks are often underestimated, including, as discussed previously, verification and validation, ongoing updates, and mitigation of AI-specific attack vectors such as data poisoning and prompt injection.

Some applications will nonetheless clear this bar, particularly in security. Cyber defence is the clearest near-term case: AI-enabled tools will almost certainly be needed for defenders to keep pace with AI-enabled adversaries. Other applications may emerge over time, but for now the most immediate risks and benefits of AI likely lie in the cyber domain.

Adapt national regulatory approaches to the speed of AI

Reflecting the "AI risk velocity" described earlier in this report, national nuclear regulatory approaches will need to adapt to an increasingly rapid pace of risk change. For now at least, this risk velocity is concentrated in the cyber and information security domains.

As Smeatham and Smeatham highlight, "the complexity of modern cybersecurity risk management requires an effective decision-making method rather than a fixed set of good practice."[51] For example, in the nuclear sector, while regulations are usually written in general terms that do not require frequent re-evaluation (and such re-evaluation is time-intensive and onerous), an "effective decision-making method" could involve regulators developing or otherwise communicating informal guidance[52] for licensees on implementing cybersecurity for nuclear facilities and activities in an AI-enabled world. That guidance can, in turn, shape the measures licensees adopt to meet regulatory requirements.

However, this path depends on regulators and licensees understanding AI and cyber defence well enough to develop high-quality guidance and implement it effectively, while the sector's technical capacity for AI is still developing. It also depends on knowledge transfer and lessons learned, which need not be purely national but could draw on international exchange and discussions with vendors. Currently, lessons from operational experience do not always flow easily between practitioners, national bodies and international standards organisations. The IAEA's international Information and Computer Security Community of Practice[53] is an important step in addressing this, and regional and national exchanges could usefully complement it.

Five Principles

This section sets out five general principles to help the nuclear sector regulate cybersecurity more effectively as AI transforms both attack and defence. Given national variation, this section does not recommend specific mechanisms or provide detailed advice, but offers considerations for the frameworks and enabling conditions that would allow a national system to build its own solution.

1. **Licensees' cost-benefit analyses for adopting AI technologies should fully account for cybersecurity, on both the cost and the benefit side.** Some cybersecurity-related costs associated with AI deployment may be routinely underestimated, including securing the AI system itself, verification and validation, ongoing updates, and mitigation of AI-specific attack vectors such as data poisoning and prompt injection. Cybersecurity-related benefits may also be underestimated, particularly for AI-enabled cyber defence tools. The benefits of adopting such tools should be weighed against the cost of failing to keep pace with AI-enabled adversaries.
2. **National regulatory bodies should develop informal guidance to help licensees implement AI systems, and seek to create settings where regulators and licensees can discuss AI and cybersecurity challenges and opportunities without licensees fearing enforcement consequences.** Informal guidance could help to address the "AI risk velocity" and, in many cases, could reference or adapt existing guidance from outside the nuclear sector. As experience accumulates, moving to draft guidance could provide more "weight" to the guidance, at the cost of some speed of adaptation. Protected, informal interactions between regulators and licensees could also help ensure that risks and expectations are well communicated in both directions. The exact mechanism for these interactions would need to be determined

51 David Smeatham and Tom Smeatham, "AI and Cyber-Resilience in the Civil Nuclear Sector", Ennuvo, May 2026.

52 The IAEA describes administrative measures — guidance, manuals, protocols, and policies — as "not legally binding" but intended to "establish the regulatory body's expectations as to how an authorised person should demonstrate compliance," and states that they "should be regularly reviewed and updated." IAEA, NSS No. 29-G.

53 More information on the IAEA Information and Computer Security Community of Practice is available online at: https://www.iaea.org/resources/webinar/webinar-on-advancing-international-collaboration-through-the-information-and-computer-security-community-of-practice.

nationally. Importantly, informal sharing is harder than it may seem, given licensees' concerns about being penalised, and this dynamic will need to be managed.

Developing and communicating such guidance will likely require regulatory bodies to build more capacity to address cybersecurity and AI. Detailed information on regulating and implementing cybersecurity in this context, provided by an authoritative international body such as the IAEA or the IEC,[54] some of which is already being developed, will be highly valuable for States that currently lack this capability. A shorter, interim document by such an international body, focused on key issues, could complement this work in the near term, helping national regulatory bodies shape their informal guidance while more detailed material is developed.

3. **Regulatory bodies should seek to build a culture of flexibility and responsiveness in cybersecurity regulatory approvals and inspections.** A stable, predictable regulatory approach and culture work well against a static or slow-changing threat, but when new adversary capabilities appear within weeks or months, the response needs to be more dynamic and may require rapidly implementing novel tools to address a shifting threat. The overall regulatory approach could gain flexibility by focusing inspections on justification and security outcomes rather than strict compliance, unless a licensee demonstrably fails to meet basic requirements. Flexibility cannot mean the absence of a standard, however: licensees need to know what they are being measured against, and it falls to the regulator to define what "good enough" looks like, even if that definition evolves over time.
4. **Vendors, licensees, and regulators should work together to share information on cybersecurity faster and more effectively.** Novel threats, such as new AI-enabled attack techniques, need to be identified and understood before they become routine. In their early stages, these cases are few, and any single organisation or country is likely to see only a handful, if any. Sharing lessons learned internationally pools these early examples, building a knowledge base fast enough to stay ahead of the threat. Valuable efforts are already underway; notably, the IAEA is developing an international Information and Computer Security Community of Practice, mentioned earlier in this report. Both national and regional communities of practice could usefully complement this work. A dedicated regional or international forum focused on innovative approaches to nuclear cybersecurity should also be considered. Participation should be voluntary, technically focused, and explicitly non-punitive; a neutral third party could organise the forum.[55]
5. **This challenge needs immediate attention.** AI-enhanced cyber threat capabilities evolve on a timescale of months, not years, increasing cyber risk over even shorter time periods if left unaddressed or only partially addressed. Yet even informal guidance of the type recommended above takes time to develop, and risk-mitigation efforts need to start immediately. Rather than waiting for perfect guidance, licensees should build capacity as effectively and quickly as possible, demonstrate that they understand the current risk landscape, set out a credible plan, and commit to addressing unknowns as they emerge, with the goal of demonstrating their competence with regulators from the outset.

In closing, discussion at the October 2026 DIET session underlined that general principles, while useful at the international level, will need to be tested against defined use cases. A natural next step would be regulatory sandbox exercises on specific applications, such as the use of AI agents for cyber defence in particular facilities and national contexts.

54 The two organisations cooperate through, for example, IEC Subcommittee SC 45A, which develops standards for the instrumentation, control, and electrical power systems of nuclear facilities, available online at: https://etech.iec.ch/issue/new-standard-for-nuclear-safety. Similar cooperation is ongoing on cybersecurity, as detailed in an IEC briefing to a 2024 IAEA conference, available online at: https://conferences.iaea.org/event/374/contributions/31132/attachments/18215/30694/ID%23292-IAEA_SMR_conf2024_IEC_cyber_security_standards_final.pdf.

55 It could draw on the model of the O-RAN ALLIANCE's Global PlugFests in the telecom sector: periodic events at which operators, vendors, and research institutions jointly test equipment against shared specifications, including security requirements, in a controlled lab environment. See O-RAN ALLIANCE, "O-RAN ALLIANCE Successfully Completes Global PlugFest Fall 2023", press release, 7 December 2023, available online at: https://www.businesswire.com/news/home/20231207289942/en.



VCDNP

Vienna Center for Disarmament and Non-Proliferation

The VCDNP is an international non-governmental organisation that promotes peace and security by conducting research, facilitating dialogue, and building capacity on nuclear non-proliferation and disarmament.

vcdnp.org

@VCDNP

info@vcdnp.org

VCDNP