Regulatory Challenges Related to the Use of Artificial Intelligence for IAEA Safeguards Verification

Cristina Siserman-Gray, Jonathan Barr, Jessica Burniske, Pegah Eftekhari, Robert Marek, Aubrey Means
Pacific Northwest National Laboratory (PNNL)

Abstract:

In the past decade, the International Atomic Energy Agency (IAEA) has recognized that artificial intelligence (AI) can provide important benefits for international nuclear safeguards through improved scope and performance, while reducing costs and manpower. But as AI is used in domains already governed by existing regulations, its uses for nuclear safeguards verifications will also have to be carefully conducted in line with regulatory requirements. The International Atomic Energy Agency (IAEA) has already been engaged in activities using AI for safeguards verification purposes and has expressed interest in further using this technology. A better understanding of the regulatory landscape for AI and how it may impact on nuclear safeguards verifications is imperative for determining whether these activities could potentially result in damage to the IAEA's institutional interests and those of its Member States. The open-source country research conducted for this paper illustrates that in majority of jurisdictions, AI legislation and regulations have just recently started to be developed and that, in several instances, AI applications may conflict with the existing frameworks governing data protection and privacy, patent and copyright laws, as well as anti-discrimination policies. Therefore, the objective of this paper is threefold: 1) provide an overview of the AI legislative and regulatory landscape in several selected States; 2) identify certain potential legal risks and challenges pertaining to the use of AI for IAEA safeguards applications; 3) provide a series of recommendations for IAEA Member States to address these challenges.

Keywords: nuclear safeguards, artificial intelligence, gaps and challenges, legal and regulatory landscape etc.

1. Introduction: The Use of AI for Safeguards Verification

Artificial Intelligence (AI) typically refers to a collection of technologies, including machine learning (ML), that process large amounts of data and algorithms with increasing computing power. AI can solve complex problems and adapt to changing patterns in ways similar to human reasoning (such as the use of inductive, deductive, and abductive reasoning),[1] and it can even possess capabilities that go beyond human reasoning, such as seeing complex patterns and signatures in data.[2] The advancement of AI presents multiple benefits, such as improving efficiency and reliability in sectors as diverse as health, the environment, and food and agriculture.

One field for which AI holds potentially great promise is the nuclear sector[3] (including nuclear science, power, and security systems), in applications staring from accelerating fusion research to optimizing the safety and security features of advanced reactor designs[4]. International nuclear safeguards, which involve technical verification measures conducted by the IAEA to ensure that States are upholding their obligations under the Treaty of the Nonproliferation of Nuclear Weapons (NPT), as well as their Comprehensive Safeguards Agreements, may also benefit from AI as it applies machine learning capabilities[5] to process the large amounts of data obtained by safeguards inspectors. This data is collected at nuclear facilities around the world through various means including video surveillance, satellite imagery, environmental samples, and mass spectroscopic analysis. To some extent, AI technologies are already being used for safeguards and nuclear security purposes, although they have not yet reached a level of full autonomy, such as the Robotic Cerenkov Viewing Device (RCVD) used to improve efficiency and accuracy of inspector verification of spent fuel rods.[6] Other types of ML systems are still in infancy, with some being developed to detect process data patterns representing the diversion of nuclear materials, or to extract important information from texts.[7] To manage and respond to the ongoing development and future deployment of AI technology, the International Atomic Energy Agency (IAEA) has established the AI for Atoms program, which convenes a series of working groups dedicated to examining the applicability of AI in the nuclear sphere.[8]

Acknowledging that AI will continue to benefit the peaceful applications of nuclear technology, the use of AI introduces legal and regulatory issues, as well as ethical and technical challenges, particularly related to data privacy and security, as well as anti-discrimination policies, including transparency, bias and trust.[9] All these risks and challenges are fundamental to considerations of the IAEA and Member States when reviewing the use of AI for safeguards verifications. The main assertion of this paper is that AI technology is not uniformly regulated across national jurisdictions globally. In fact, as will be demonstrated in the following sections, while in some jurisdictions the legislative and regulatory landscape for the use of AI technology is still developing, in others, the use of this technology is subject to multiple rules in ways different from traditional information systems, including laws and regulations on data protection and privacy, patent and copyright laws, as well as anti-discrimination policies.[10] These legal frameworks may impact the way IAEA safeguards data can be collected, received, or analyzed. Consequently, the first part of this paper will provide an analysis on the level of regulation of AI in a number of Nuclear Weapons States (NWS) and Non-Nuclear Weapons States (NNWS). In the second part, the paper will highlight examples of current AI applications for safeguards verification purposes and identify potential legal risks for the IAEA and Member States pertaining to this use. Finally, the paper will conclude by discussing how the AI legal framework of IAEA Member States may need to be adjusted before proceeding with a wider use of AI technologies for safeguards verification purposes.[11]

2. National Legal Frameworks on Artificial Intelligence

Currently, there is no global framework for regulating AI, and different countries have taken different approaches to regulating it. Beyond a few international initiatives, including "*AI for Good*", *"Global Partnership on AI"*, and "*OECD Principles on AI*", which have been engaged in promoting the responsible development if AI, the efforts in this space have been limited. There are several reasons why that is the case. First, AI is a rapidly developing technology, and it is difficult to keep up with the pace of change; and second, AI is a complex technology, for which it is difficult to identify potential risks. While there is no consensus on the best way to regulate AI, the main challenge is ascertaining whether existing laws can regulate it in conformity with established legal, moral, and ethical principles and, if not, what new legal instruments are necessary to meet that objective.[12] To help illustrate current approaches to regulating AI, the relevant regulatory framework in several IAEA Member States, both Nuclear Weapons States (NWS) and Non-Nuclear Weapons States (NNWS), will be briefly introduced below. Having a better understanding on the current AI regulatory frameworks in these countries informs the larger conversation on how IAEA may respond to a wider use of AI based applications for safeguards verification purposes.

2.1. AI Regulations in Nuclear Weapons States

The aspects of AI that are regulated vary from country to country, but common aspects include data privacy, discrimination, safety of AI systems and accountability. Most NWS have AI-relevant laws that address the apportionment of liability for injuries resulting from unreasonable behaviors or defective products that use AI technologies, that define intellectual property rights, and that seek to ensure fairness in decisions and protect privacy.[13] Many NWS have started to consider the use of AI for a number of nuclear applications, including nuclear reactors design, fuel cycle management, as well as for nuclear security and safety purposes. However, most NWS still encounter challenges in applying traditional rules to AI, particularly as concerns the nuclear field, and regulatory agencies and legislatures must determine whether special rules are needed for addressing AI for safeguards verifications, among other uses and concerns.[14]

The United States (U.S.) recently released its *Blueprint for an AI Bill of Rights* (the "Blueprint"), issued by the White House Office of Science and Technology Policy.[15] The Blueprint is guided by several key principles, including algorithmic discrimination protections and data privacy. The Blueprint calls upon "designers, developers, and deployers of automated systems [to] take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way."[16] It also discusses the appropriate use of surveillance and monitoring technologies and situations where such technologies should not be used (e.g., education, work, housing, "or in other contexts where the use of such surveillance technologies is likely to limit rights, opportunities, or access").[17] Finally, the Blueprint calls for the opportunity for individuals to opt out of automated systems in favor of human decision-making.

AI-relevant provisions have been included in certain aspects of U.S. legislation, but there has not been comprehensive legislation to address the use of AI in the public or private sector. At the state level, several jurisdictions have adopted data privacy laws or government initiatives aimed at AI that seek to address emerging challenges from the use of those technologies,[18] but efforts to regulate AI are piecemeal and have not been coordinated at the federal level. Publicly available information does not indicate whether the United States is currently using AI to conduct safeguards verification activities, although the U.S. government and associated entities have previously carried out research in this area.[19] The U.S. government is reportedly seeking to use AI in nuclear power plants for other applications, including to prevent cyberattacks.[20] A recent report by the U.S. Nuclear Regulatory Commission also examined the use of AI/ML in nuclear power plants,[21] identifying ways that the technology could be applied. While the report did not mention safeguards verification, it did examine the potential applications of AI/ML in areas such as plant safety and security, plant operation and maintenance, and accident diagnosis and prognosis.[22]

In other NWS, progress on AI has been more pronounced. In France, the government has adopted a *National Strategy on AI for 2018-2026*, which sets as its objectives attracting and investing in AI, disseminating AI and big data in the economy, and promoting an ethical AI system. In 2019, the

National Ethics Committee established a Digital Ethics Pilot Committee, aiming to address ethical issues of digital tools and AI.[23] As a member of the European Union (EU), France is also following the EU's framework on data collection, which is regulated by the EU's Regulation 2016/679 or the General Data Protection Regulation (GDPR). The country has already started looking at the implications of AI for the nuclear field. For example, the *AI Research on Data for Nuclear Application (ARDNA)* project supports investment and modernization of the nuclear industry as part of the initiative "France Relance."[24] Although open-source research did not reveal any use of AI for purposes of safeguards verification in France, it is expected that the AI efforts currently conducted in the country may open the door for the use of AI technologies for safeguards verification.

The United Kingdom's (U.K.) *National AI Strategy* (the "Strategy") explicitly acknowledges that the U.K. has not yet adopted laws explicitly to regulate AI, although it points to a "patchwork of legal and regulatory requirements built for other purposes which now also capture uses of AI technologies."[25] For instance, the current U.K.'s *Data Protection Act* includes specific requirements around automated decision-making and the processing of personal data, which "also covers processing for the purpose of developing and training AI technologies."[26] The U.K.'s *2022 Defense Artificial Intelligence Strategy* also states that the U.K. will "study the effects of AI on the inter-linked domains of cyber, space, and nuclear, examining AI's potential to accelerate or amplify developments linked to other emerging and strategic technologies."[27] While the U.K. uses AI for certain tasks at its Sellafield nuclear facility,[28] none of these tasks appear related to nuclear safeguards, but with UK's Defense Artificial Intelligence Strategy in place, this could, however, open the possibility for such usage in the near future.

Lastly, China's 2017 national strategy on AI, labeled "Notice of the State Council on Printing and Distributing the Development Plan for a New Generation of Artificial Intelligence,"[29] prioritizes the responsible development of legal and regulatory norms related to AI. China's *Internet Information Service Algorithmic Management* (IISARM) *Regulations* came into effect on March 1, 2022[30] and prevents service providers from engaging in activities endangering national security and social public interests, although these concepts are not clearly defined in the law. Chinese private companies have already developed or purchased AI software intended to be used as surveillance devices, capable of analyzing large amounts of data. Although the software is currently utilized by law enforcement, there seems to be a large-scale effort in China to upgrade their technological capabilities by tapping into the power of big data and AI. In this regard, China has issued a three-year implementation plan for Internet+ AI, putting forward a series of measures for technology R&D, application and industrial development. While open-source literature is silent on the use of AI for nuclear safeguards verifications in China, it is expected that the country will continue exploring the application of AI for these purposes.

2.2. AI for Safeguards Verifications in Non-Nuclear Weapons States

The following sections highlight how certain NNWS have approached the use and regulation of AI, including laws, regulations, and policies that are being considered or have been implemented, and any specific safeguards applications of AI. This section highlights a select number of EU countries, as well as a few other countries around the world, which have recently made a significant progress in this area. Due to the limited scope of this research, the legal mapping below provides just a glimpse on the complex gaps related to the use of AI and a high-level picture of the IAEA's Member States regulatory efforts in this area.

2.2.1 European Union

European Union is one of the major players in the global AI race and has invested heavily in AI research and development. In the regulatory space, the European Union is currently in the process of developing its *Artificial Intelligence Act*, which is a regulation proposed in April 2021 by the European Commission with the purpose to introduce a common regulatory and legal framework for AI[31]. The Act regulates the providers of AI and entities making use of these systems in a commercial capacity. The EU has also developed an *AI Strategy*, which articulates several key principles related to the use of AI systems.[32] The Strategy aims at making the EU a world-class hub for AI and ensuring that AI is human-centric and trustworthy.

One area where the EU has been highly influential involves its adoption of the *General Data Protection Regulation (GDPR)* in 2016, which contains several provisions relevant to automated decision-making (while distinct from AI)[33] and the transfer of personal data to third countries or international organizations. GDPR requires companies engaged in big data, machine learning and AI to ensure that: a) processing of personal data during AI phases follows specified, explicit and legitimate purpose; b) processing of data has a legal basis as listed in GDPR, Article 6; c) the data is stored for a limited and specified time;[34] d) data is collected only when it is strictly necessary; e) no data is transferred outside the EU; f) principles of privacy by design and privacy by default are respected; and g) personal data is appropriately secured. With respect to automated decision-making, GDPR establishes that data subjects "shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significant affects him or her."[35] A study conducted by the EU Parliament on the impact of GDPR on AI concluded that the two are compatible[36]. While acknowledging that there is a tension between the traditional data protection principles and the fully deployment of AI, the study concludes that "there are ways to interpret, apply and develop data protection principles that are consistent with the beneficial uses of AI".

Within EU, several countries have adopted initiatives on AI, which go beyond the EU framework discussed above. For instance, Finland, through its Ministry of Employment and Economy, has established a steering group on implementation of the elements of AI to enable lifelong learning and support regarding digital skills development and free AI courses for all its citizens.[37] The country also has put in place an action plan to speed up the introduction of AI technologies and promote the fourth industrial revolution in the country. While Finland does not yet appear to have adopted or introduced AI-specific regulations specific for nuclear industry, nor revealed information regarding the use of AI in its safeguards verification program, given its strong initiatives to support the use of AI, the country is expected to promote this use.

In a similar effort, Italy has formed a group of experts to identify a *National Strategy for Artificial Intelligence,* published in July 2020.[38] In November 2021, the Ministries of Education, University and Research, Economic Development, and Technological Innovation and Digital Transition produced Italy's *Strategic Program for AI 2022-2024*.[39] This document, developed in line with the European

Commission's Strategy on AI,[40] fully embraces the adoption of AI and its applications in public administration, including the energy sector. As part of its guiding principles, Italy commits to a "trustworthy and sustainable" AI that revolves around principles of responsible use of data and AI technology. Although Italy has not adopted AI-specific laws and regulations,[41] its Council of State has clarified that fully automated systems should comply with principles of transparency regarding the process of decision-making mechanisms, and ensuring non-discrimination.[42] The open-source research did not reveal any information regarding the country's position on the use of AI for safeguards verification programs, but identified areas where AI is already used by the nuclear sector.

2.2.2. Canada

Canada's Minister of Innovation, Science and Industry sponsored in 2022, Bill C-27, which among other acts proposes the enactment of the *Artificial Intelligence and Data Act* (the "AIDA"), Canada's first attempt to formally regulate certain AI systems, apart from its other privacy legislations. If passed, AIDA would regulate the design, development, and use of certain types of AI systems and ensure that developers and operators of such systems adopt measures to mitigate various risks of harm and avoid biased outputs. AIDA also mandates impact assessments for AI systems and subjects "high-impact" systems to further public disclosure requirements, among others. AIDA also establishes prohibitions related to the possession or use of illegally obtained information for the purpose of designing, developing, using or making available for use an AI system if its use causes serious harm to individuals or national interests.[43] Specific to the nuclear field, Canada's Nuclear Safety Commission (CNSC) noted that it was working with nuclear consulting firms to review the utility of AI in support of CNSC's regulatory framework and its readiness to regulate the use of AI. As part of its assessments, CNSC will review the existing international regulatory activities surrounding AI applications in nuclear and provide a report by the end of 2023 to all stakeholders.[44]

2.2.3 Japan

Japan's Cabinet Office, in its 2022 AI Strategy, sets out objectives for utilizing AI to enhance Japan's industrial competitiveness.[45] In January 2022, the Ministry of Economy, Trade and Industry (METI) issued the Governance Guidelines for *Implementation of AI Principles, Ver. 1.1*,[46] which provide action targets, practical examples and gap analysis. The guidelines are expected to support companies engaged in AI development and operation systems to voluntarily implement AI social principles, adhere to a human-centric approach to AI, and provide for privacy protections, security, and accountability and transparency. In addition, the National Institute of Advanced Industrial Science and Technology (AIST) published the *Guidelines on Machine Learning Quality Management*, which address topics such as vulnerabilities of machine learning. For nuclear industry, it is unclear whether the country is contemplating uses of AI for safeguards verification activities. However, for purposes of Units 1 and 3 of the Fukushima Daiichi Nuclear Power Station (NPS) where normal inspections cannot be carried out due to inaccessibility of the reactors, Japan's Nuclear Regulatory Authority (NRA) has introduced a full-time monitoring system based on the use of surveillance cameras and radiation monitors and special additional verification activities. In addition, in FY 2021, the NRA entered into an agreement with the IAEA regarding new inspection procedures, similar to the approach employed at the Tokyo Electric Power Company (TEPCO's) Fukushima Daiichi Nuclear Power Station (NPS), including other research and development facilities, research reactors, and criticality facilities. Thus, it is foreseeable that in the near future AI may be utilized to process the data collected from this full-time monitoring system.

The regulatory mapping provided in this section illustrates that for the foreseeable future, in both NWS and NNWS, the initiatives for the regulation of AI will most probably continue to be highly context-dependent, varying considerably on what objectives the regulatory scheme is intended to achieve. That is because regulating AI for the purpose of data protection at an individual level, for example, is different from broader regimes regulating AI as part of wider policies affecting entire organizations and States. Therefore, a more detailed legal analyses will be necessary in order to identify common trends and discrepancies among the various ways IAEA Member States regulate AI in their national frameworks in general, but also in the nuclear energy and national security field, more specifically. A more detailed analyses would be informative to the IAEA, not only because they would provide a better understanding of the potential contentious areas of AI, but also would be indicative of those fields in which IAEA might have to adapt its policy frameworks to ensure that Member States will agree on the use of AI-systems to draw conclusions on safeguards verifications activities.

3. Challenges and Opportunities for AI in Safeguards Verification Purposes

Safeguards involve large amounts of heterogeneous data, dynamic data sets, and highly complex operations, which require significant investment of resources to analyze effectively and efficiently. Because of this, international nuclear safeguards are under constant pressure to be more efficient to cope with an expansion of global nuclear fuel cycle activities, increasing nuclear proliferation threats, and a constrained budget for the IAEA. However, the IAEA anticipates that AI could have both positive and negative effects on systems, processes, and procedures relevant to international nuclear safeguards. For instance, AI could significantly improve safeguards efficiency by focusing on value-added tasks and reducing unnecessarily repetitive ones. At the same time, these technologies may introduce new sources of uncertainty and reduced transparency. Potential positive and negative impacts of these systems have not yet been well understood. This section will examine a series of applications in which IAEA safeguards inspectors might use AI technologies and provide an analysis on how these applications may be hindered by the difficulty of demonstrating compliance with the regulatory standards.

3.1. AI for Verification of Spent Fuel

Spent fuel is measured by utilizing the neutrons and gamma rays emitted and can also be performed using Cerenkov imaging data. These inspections generate a large amount of data. The data sets may be used to build AI algorithms, where numerical simulations can supply training and test datasets for the model.

Several AI models have been examined to determine if models can successfully distinguish between complete fuel assemblies and defective fuel assemblies. AI has also been used to improve the processing of data obtained from the next Generation Cerenkov Viewing Device (XCVD) via a support vector machine to classify blurry XCVD images, which may require further image processing. For example, the Robotic Cerenkov Viewing Device (RCVD) has been created through a collaboration between the Australian national science agency's (CSIRO) data and digital specialist arm Data61, Hungarian robotics company Datastart, and the IAEA.[47]

Although AI/ML has the potential to efficiently monitor spent fuel assemblies, guaranteeing the accuracy of the technology is critical to making its application effective. A 2020 study on the use of AI in the nuclear industry found that the behavior or accuracy of intelligent systems for dynamic situations, such as the innate characteristics of a nuclear reactor, are vulnerable to fooling and may produce inaccurate results. Multiple studies[48] have demonstrated the potential to fool deep neural network models in the classification of unrecognizable images, and that changing the orientation of an object can lead to it being mislabeled by the technology.[49] The "black box" nature of AI, where the knowledge is virtually baked into the technology and cannot necessarily be explained by operators, can make AI's rationale difficult to discern or alter. Any AI technology applied to spent fuel verification, therefore, must be rigorously tested and sufficiently advanced to ensure accuracy.

Therefore, assurances that AI-based systems perform as intended are necessary, as the quality, accuracy, and relevance of data are essential for safeguards verification purposes. Any data bias, error, or statistical distortion will be learned and amplified. In situations involving ML—where algorithms and decision rules are trained using data to recognize patterns and to learn to make future decisions based on these observations—regulators and users may not easily discern the properties of these algorithms. These algorithms are able to train systems to perform certain tasks at levels that may exceed human ability. Therefore, they raise many challenging questions, including calls for greater algorithmic transparency to minimize the risk of bias, discrimination, unfairness, and error to ultimately protect Member States' interests.

3.2. AI and Satellite Imagery

The importance of satellite imagery goes beyond simply verifying States' declarations, planning and supporting verification activities, and detecting and investigating undeclared activities. Satellite imagery also plays a significant role in monitoring nuclear fuel cycle activities.[50] With this in mind, an extensive range of safeguards data collected from satellites could be used to train ML algorithms, particularly given the growing volume of available satellite imagery and open-source data.

Commercial satellite imagery has become a very important information source for the IAEA's Department of Safeguards, especially in places where the IAEA does not have physical access.[51] For instance, satellite imagery helps the IAEA keep abreast of developments in the nuclear program of the Democratic People's Republic of Korea (DPRK), even though it is unable to carry out physical verification activities there. Monitoring developments at the Yongbyon plutonium production site is particularly important. The use of satellite imagery has allowed the IAEA to prepare and update a detailed plan for the implementation of monitoring and verification activities in the DPRK in the event of inspectors returning to that country.

However, issues related to transparency, scope, and privacy are often cited when discussing the use of satellite imagery, and the same is especially true with the involvement of AI. As the RAND Corporation has reviewed in a series of workshops on the nuclear security implications of AI over the next century,[52] merely the perception of AI as potentially destabilizing could raise issues with potential implications on national security. For example, if a State under a comprehensive safeguards agreement believed that the IAEA, through its AI capabilities, had access to or was using satellite imagery to observe more than just the declared nuclear facilities subject to safeguards, the State could fear a violation of its sovereignty and a risk to its national security interests. Additionally, as with any data-based system, there is a concern related to control of the data and who has access to it. A State may conduct a safeguards agreement with the IAEA, but what controls are in place to ensure that the AI-gathered satellite data is not obtained by a malicious state or non-state actor? Therefore, for AI use of satellite imagery to be secure, both the scope of the data obtained and exclusive IAEA control over that data should be clearly defined. These aspects need to be clearly reflected in the regulatory frameworks of the Member States.

3.3 AI and Video Surveillance

Video cameras have long been relied upon by the IAEA, along with various sensor technologies, to generate a complex and growing amount of data which can be used in various ways. For example, in 2021 the IAEA maintained over 1,300 surveillance cameras at nuclear facilities around the world that operated around the clock to ensure continuity of knowledge of nuclear material and installations. All of that data was then collected and reviewed by safeguards inspectors to verify authorized control of nuclear material.[53] Another area of recent research was the review of surveillance data to detect and track safeguards-relevant objects, operator declarations, and anomalous activities. Surveillance data has also been used with learning-based algorithms to detect and count objects within a nuclear facility.

An ongoing challenge given the limited resources at the IAEA is the amount of effort needed to annotate training datasets. The ability to more efficiently analyze video surveillance data using AI would hold significant benefits for international safeguards by reducing cost and workforce effort. As the demands on IAEA safeguards inspectors grow with the number of nuclear facilities around the world, the IAEA budget remains static; maximizing efficiency of safeguards operations and improving inspector output is therefore a high priority that AI may help to address. In this respect, AI could help automate the annotation process as well as reduce the overall amount of training data required. This would enable safeguards verifications to be conducted more efficiently.

However, before using these technologies, further improvements to AI implementation are needed to better train the models. Also, similar to the concerns raised by the use of satellite imagery, control of data used in video surveillance is particularly sensitive for States' security concerns. The advance of facial recognition technology and geolocation tools may create vulnerabilities for Member States under safeguards if such data is obtained by malicious actors. States that enter into safeguards agreements with the IAEA must be assured that the data collected by AI does not fall into the wrong hands.

4. Recommendations and Conclusions

The emergence of AI is revolutionizing the nuclear energy sector, and its use in safeguards verification is no exception. As the IAEA and Member States continue to seek more efficient, cost-effective, and safe ways to perform their safeguards verifications, AI could have a major impact. While the AI legal regime is developing in several countries throughout the world, the AI technology proposed to be used for safeguards verifications still needs to be developed in line with existing legal, regulatory, and policy requirements.[5455]

Some legal and regulatory areas that would require immediate attention are:

- **Intellectual property (copyright and patents)**: While intellectual property law is continuing to take shape around AI, legal authorities have recently emphasized that AI cannot function as a "person" under copyright and patent law. This will become an area of interest for both IAEA and Member States as more AI technologies for safeguards verifications are developed.
- **Anti-discrimination policies**: As algorithms and AI-based systems make decisions affecting individuals, a growing concern has been raised about whether such decisions are fair and reliable. This is particularly relevant for safeguards verification as it goes beyond individuals and impacts organizations and States interests at large.
- **Data protection, reliability, and transparency**: In the case of data protection infringement, enforcement possibilities include a reprimand or a temporary or definitive ban on processing. In some countries, regulatory authorities may also be subject to administrative fines. This is another area relevant for safeguards verification, particularly as government authorities with safeguards roles collect and analyze data.
- **Privacy and surveillance:** Concerns around AI/ML-based surveillance and data protection have become stronger. As more AI/ML systems will be used for verifications, these systems will raise a host of concerns on transparency, explainability, fairness, privacy, and accountability.

Given that nuclear and AI disciplines contain a certain degree of risk and uncertainty, it is advisable that the IAEA continues to be an active participant in the regulatory developments in this field. A better understanding of the regulatory landscape for AI and how it may impact on nuclear safeguards verifications is imperative for determining whether these activities could potentially result in damage to the IAEA's institutional interests and those of its Member States. Therefore, IAEA should engage with Member States by: 1) raising awareness among practitioners of the legal and ethical impacts of AI technology on nuclear science and applications; 2) creating mechanisms for dialogue among stakeholders; and 3) establishing responsible governance of AI technology applications in the nuclear field.[56]

Lastly, while AI may not replace safeguards inspectors anytime soon, the current interest of Member States to regulate AI suggests that AI-based technologies will become more used for a variety of applications, including safeguards inspections. Because the future of AI is open-source, the IAEA may need to adapt its policy frameworks to provide this data. While some safeguards data is open-source, there is a portion of data IAEA receives and securely analyzes. This structure may need to be adapted to allow for specific instances of providing data. IAEA's legal framework and those of its Member States may also need to be altered to enable data sharing. Nonetheless, although the threat of hacking is not specific to AI, as long as AI involves computers then the possibility of data tampering must be considered – particularly when data is related to nuclear facilities or material. Sufficient security measures should be in place to retain control over AI data, to prevent both its manipulation and leakage to parties outside of the IAEA. These are all areas that warrant additional attention particularly as concern the legal and regulatory issues they raise

Endnotes

[1] AJ Abdallat, *Why Human-Like Reasoning is the Key to Trusted AI*, Forbes (Aug. 30, 2017), https://www.forbes.com/sites/forbestechcouncil/2017/08/30/why-human-like-reasoning-is-the-key-to-trusted-ai/?sh=739df4af4edf.

[2] Broussard, E., *The Future of Atoms: Artificial Intelligence for Nuclear Applications*, September 23, 2020, https://www.iaea.org/newscenter/news/the-future-of-atoms-artificial-intelligence-for-nuclear-applications, accessed April 24, 2023.

[3] Wastin, F., et al., *Horizon Scanning for Nuclear Safety, Security and Safeguards: Yearly Report 2019*, European Commission, 2020.

[4] Boehnlein, A. et al, *Artificial Intelligence and Machine Learning in Nuclear Physics*, 2021, https://www.researchgate.net/publication/356817844_Artificial_Intelligence_and_Machine_Learning_in_Nuclear_Physics/citation/download, accessed April 24, 2023.

[5] Wehsener, A., et al., *Forecasting the AI and Nuclear Landscape*, Institute for Security and Technology, 2022.

[6] Australian Technology Helping Safeguard Used Nuclear Fuel - Nuclear Engineering International, https://www.neimagazine.com/news/newsaustralian-technology-helping-safeguard-used-nuclear-fuel-10548131, accessed April 24, 2023.

[7] PNNL, *Detecting Nuclear Threats with Artificial Reasoning*, 2022, https://www.pnnl.gov/news-media/detecting-nuclear-threats-artificial-reasoning, accessed April 24, 2023.

[8] see https://nucleus.iaea.org/sites/ai4atoms/SitePages/Home.aspx (Accessed May 11, 2023).

[9] I. Chiu and E. Lim, *Managing Corporations' Risk in Adopting Artificial Intelligence: A Corporate Responsibility Paradigm*, Washington University Global Studies Legal Review, Vol. 19, Issue 347, 2021.

[10] AI technology presents unique challenges to privacy because it can expose information differently from traditional information systems, and it can do so in unpredictable ways. Users of AI systems like ChatGPT have highlighted unique ways of posing questions that can expose information that the AI system would otherwise be prevented from providing to the user. *See*, for instance, https://www.reddit.com/r/ChatGPT/comments/12dt535/reverse_psychology_always_works/ (accessed May 10, 2023).

[11] International Atomic Energy Agency, *Artificial Intelligence for Accelerating Nuclear Applications, Science and Technology, Artificial Intelligence for Accelerating Nuclear Applications*, Science and Technology, 2022. https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology, accessed April 24, 2023.

[12] Cath, C., *Governing artificial intelligence: ethical, legal and technical opportunities and Challenges*, Royal Society Publishing, 2018, https://royalsocietypublishing.org/doi/10.1098/rsta.2018.0080, accessed April 24, 2023.

[13] Schildkraut, P., *AI Regulation: What You Need to Know to Stay Ahead of the Curve*, 2021, https://www.arnoldporter.com/-/media/files/perspectives/publications/2021/06/ai-regulationstaying-ahead-of-curveschildkraut0621.pdf?la=en&rev=8b51c39301374e5fb429f29303af9918, accessed April 24, 2023.

[14] Congressional Research Service, *Artificial Intelligence and National Security*, 2020, https://sgp.fas.org/crs/natsec/R45178.pdf, accessed April 24, 2023.

[15] The White House, Blueprint for an AI Bill of Rights, https://www.whitehouse.gov/ostp/ai-bill-of-rights/, accessed April 24, 2023.

[16] Id.

[17] Id.

[18] *see, e.g.*, California, Utah, Colorado, Virginia, and Connecticut, states that have each adopted data privacy laws. The California Consumer Privacy Act and California Privacy Rights Act have proven to be particularly influential in this regard.

[19] DOE, *NNSA Leads National Collaboration to Drive Next-Generation in AI for Nonproliferation*, 2021. https://www.energy.gov/nnsa/articles/nnsa-leads-national-collaboration-drive-next-generation-ai-nonproliferation, accessed April 24, 2023.

[20] Keller, J., *U.S. Nuclear Regulators Seek to Apply AI and Machine Learning to Cyber Security at Nuclear Power Plants*, Military Aerospace Electronics, 2022. https://www.militaryaerospace.com/trusted-computing/article/14233131/cyber-security-nuclear-power-ai-and-machine-learning, accessed April 24, 2023.

[21] *See also* Department of Energy, *FAIR data and models for artificial intelligence and machine learning*, 2020, https://science.osti.gov/-/media/grants/pdf/lab-announcements/2020/LAB_20-2306.pdf, accessed April 24, 2023.

[22] Ma, et al., *Exploring Advanced Computational Tools and Techniques with Artificial Intelligence and Machine Learning in Operating Nuclear Plants*, U.S. Nuclear Regulatory Commission, 2022). https://www.nrc.gov/docs/ML2204/ML22042A662.pdf, accessed April 24, 2023.

[23] Global Legal Insights, AI, Machine Learning and Big Data Laws and Regulations, 2022. https://www.globallegalinsights.com/practice-areas/ai-machine-learning-and-big-data-laws-and-regulations/france#chaptercontent5, accessed April 24, 2023.

[24] Gerlat, P.*, France Relance: ARDNA, AI project winner of the call for projects to support investment and modernization of the nuclear industry*, March 2022. https://www.actuia.com/english/france-relance-ardna-ai-project-winner-of-the-call-for-projects-to-support-investment-and-modernization-of-the-nuclear-industry/, accessed April 24, 2023.

[25] United Kingdom, Guidance: National AI Strategy, September 2021, https://www.gov.uk/government/publications/national-ai-strategy, accessed April 24, 2023.

[26] Id.

[27] Id.

[28] United Kingdom, Guidance: National AI Strategy, September 2021. https://www.gov.uk/government/publications/sellafield-ltd-ai-strategy, accessed April 24, 2023. *See also* . World Nuclear News, Sellafield AU Strategy to Boost Safety and Speed Up Site Remediation, March 2023. https://www.world-nuclear-news.org/Articles/Sellafield-unveils-AI-strategy-to-accelerate-clean, accessed April 24, 2023.

[29] The State Council on printing and distributing Notice of the New Generation Artificial Intelligence Development Plan, 2017. http://www.gov.cn/zhengce/content/2017-07/20/content_5211996.htm, accessed April 24, 2023.

[30] Cyber Administration of China, Provisions on the Administration of Internet Information Service Algorithm, 2022. http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm, accessed April 24, 2023.

[31] European Commission, *Regulation of the EU Parliament and the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts*, Brussels, 2021.

[32] Communication for the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions, *Artificial Intelligence for Europe*, Brussels, 2018.

[33] Autonomous systems can be thought of as completing tasks, while AI engages in problem-solving to complete tasks that would normally require human intelligence. *See* https://inertialsense.com/artificial-intelligence-vs-autonomy (accessed May 8, 2023).

[34] Technologies such as reverse engineering could make it possible for AI models to recreate the data, however.

[35] Id. at Article 22.

[36] European Parliament, The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence, June 2020, https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf, accessed April 24, 2023.

[37] Finland, *Steering Group on AI*, 2021. https://tem.fi/hanke?tunnus=TEM104:00/2021, accessed April 24, 2023.

[38] Italy, Ministry of Industry, *AI Strategy*, 2020. www.mise.gov.it/index.php/it/198-notizie-stampa/2041246-intelligenza-artificiale-online-la-strategia, accessed April 24, 2023.

[39] Italy, *Strategic Program on Artificial Intelligence*, Rome, 2021. https://assets.innovazione.gov.it/1637777513-strategic-program-aiweb.pdf, accessed April 24, 2023.

[40] Regulation of the EU Parliament and the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, Brussels, 2021.

[41] *But see* https://www.bbc.com/news/technology-65139406 (accessed May 5, 2023), describing how Italy has blocked ChatGPT due to privacy concerns, although not due to any AI-specific laws.

[42] Consiglio di Stato sez. VI, 13/12/2019, Decision no. 8472.

[43] Fasken, *The Regulation of Artificial Intelligence in Canada and Abroad: Comparing the Proposed AIDA and EU AI Act*, October 2022. https://www.fasken.com/en/knowledge/2022/10/18-the-regulation-of-artificial-intelligence-in-canada-and-abroad, accessed April 24, 2023.See also Medeiros, M., Beatson, J., *Data Protection Report*, Norton Rose Fulbright, 2022. https://www.dataprotectionreport.com/2022/06/canadas-artificial-intelligence-legislation-is-here//, accessed April 24, 2023.

[44] Canadian Nuclear Safety Commission, *Remarks by Rumina Velshi at the Canadian Nuclear Society's 3rd Annual International Conference on Disruptive, Innovative and Emerging Technologies in the Nuclear Industry*, November 2022. https://www.canada.ca/en/nuclear-safety-commission/news/2022/11/remarks-by-rumina-velshi-at-the-canadian-nuclear-societys-3rd-annual-international-conference-on-disruptive-innovative-and-emerging-technologies-in.html, accessed April 24, 2023.

[45] Japan Cabinet Office, *AI Strategy*, 2022. https://www8.cao.go.jp/cstp/ai/index.html, accessed April 24, 2023.

[46] Ministry of Economy, Trade and Industry (METI), *Governance Guidelines for Implementation of AI Principles*, January 2022. https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20220128_2.pdf, accessed April 24, 2023.

[47] World Nuclear News, *Robot developed to assist verification of used fuel*, 2023. https://www.world-nuclear-news.org/Articles/Robot-developed-to-assist-verification-of-used-fue, accessed April 24, 2023.

[48] Nguyen, A., Yosinski, J., Clune, J., *Deep neural networks are easily fooled: high confidence predictions for unrecognizable images*, IEEE Conference on Computer Vission Pattern Recognition, 2015. *See also* M.A. Alcorn, et al., *Strike (With) a Pose: Neural Networks Are Easily Fooled by Strange Poses of Familiar Objects,* 2018.

[49] Suman, S., Artificial Intelligence in Nuclear Industry: Chimera or Solution? *Journal of Cleaner Production* Vol. 278, 2021.

[50] Sein, G. et al., *International Safeguards and Satellite Imagery*, Springer, 2009. https://link.springer.com/book/10.1007/978-3-540-79132-4, accessed April 24, 2023.

[51] Quevenco, R., *Completing the picture: using satellite imagery to enhance IAEA safeguards capabilities*, 2016. https://www.iaea.org/publications/magazines/bulletin/57-2/completing-the-picture-using-satellite-imagery-to-enhance-iaea-safeguards-capabilities, accessed April 24, 2023.

[52] Geist, E. et al, *How Might Artificial Intelligence Affect the Risk of Nuclear War*?, 2018. https://www.rand.org/pubs/perspectives/PE296.html, accessed April 24, 2023.

[53] Wagman, J., *The Evolution of Safeguards Technology*, 2022. https://www.iaea.org/bulletin/the-evolution-of-safeguards-technology, accessed April 24, 2023.

[54] Ruschemeier, H., *AI as a challenge for legal regulation*, ERA Forum, Vol. 23, Issue 3, p. 361-376, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9827441/, accessed April 24, 2023.

[55] Henz, P., *Ethical and legal responsibility for Artificial Intelligence*, Discover Artificial Intelligence, Vol. 2, 2021, https://link.springer.com/article/10.1007/s44163-021-00002-4, accessed April 24, 2023.

[56] IAEA, *Artificial Intelligence for Accelerating Nuclear Applications*, Science and Technology, 2022. https://www.iaea.org/publications/15198/artificial-intelligence-for-accelerating-nuclear-applications-science-and-technology, accessed April 24, 2023.