Journal of International Technology and Information Management

Volume 31 | Issue 3 Article 4

2023

The Need for International AI Activities Monitoring

Parviz Partow-Navid
California State University, Los Angeles, ppartow@calstatela.edu

Ludwig Slusky
California State University, Los Angeles, lslusky@calstatela.edu

Recommended Citation

Partow-Navid, Parviz and Slusky, Ludwig (2023) "The Need for International AI Activities Monitoring," *Journal of International Technology and Information Management*: Vol. 31: Iss. 3, Article 4.
DOI: https://doi.org/10.58729/1941-6679.1564
Available at: https://scholarworks.lib.csusb.edu/jitim/vol31/iss3/4

The Need for International AI Activities Monitoring

Parviz Partow-Navid
(*California State University, Los Angeles*)

Ludwig Slusky
(*California State University, Los Angeles*)

ABSTRACT

This paper focuses primarily on the need to monitor the risks arising from the dual-use of Artificial Intelligence (AI). Dual-use AI technology capability makes it applicable for defense systems and consequently may pose significant security risks, both intentional and unintentional, with the national and international scope of effects. While domestic use of AI remains the prerogative of individual countries, the unregulated and nonmonitored use of AI with international implications presents a specific concern. An international organization tasked with monitoring potential threats of AI activities could help defuse AI-associated risks and promote global cooperation in developing and deploying AI technology. The paper reviews factors involved in the international monitoring of AI activities, strategies of dual-use technologies regulation, challenges, and potential solutions.

Keywords: AI, International, Monitoring

INTRODUCTION

Artificial Intelligence (AI) is fast becoming the most transformative technology in the world, and its potential to revolutionize every aspect of human life is unprecedented. This technology is already playing a critical role in shaping the world's economy, defense, and security systems. However, with great power comes great responsibility, and the development of AI technology raises concerns about its potential misuse, particularly in the area of national security. There is a growing need for an organization like the United Nations (UN) Atomic Nuclear Agency to monitor countries' activities in AI development and deployment to mitigate the risks associated with the misuse of AI technology.

Researchers (Seger et al., 2023) emphasize that overseeing AI developments at the international level will promote the democratization of AI technology which would restrict reckless decision-making in fierce competition for "benefits of choices around AI use, development, and distribution of profits." Such *democratization of*

ISSN: 1941-6679-On-line Copy

AI governance under the international umbrella will help to develop complex normative decisions and methods to involve public participation and diverse and substantive representation of stakeholder viewpoints. Significantly, it could also facilitate "well-informed deliberation, … fair and open election processes, [and] constitutional protections for individuals and minorities."

Some of the other aspects that require careful studies include privacy protection, which could be one of the most complained casualties of AI proliferation; the high cost of training AI models, which makes it prohibitively expensive for smaller organizations and individuals, thus limiting competitive differentiation; and merging robotics with AI (e.g., for law enforcement).

This research aims to examine the challenges involved in *establishing an international organization to monitor AI activities* and propose potential solutions that address these challenges. The research focuses on understanding the need for consensus on responsible AI development, the complexities of monitoring rapidly evolving AI technologies, the allocation of financial resources, and the concerns about the potential misuse of monitoring powers. The aim is to provide practical recommendations and insights for establishing and operating an AI monitoring organization that promotes global cooperation, transparency, and accountability in developing and using AI technology.

The authors arrived at this manuscript by realizing the increasing significance and transformative nature of Artificial Intelligence (AI) in today's world. They appreciate the potential of AI to alter various aspects of human life but also acknowledge the concerns regarding its misuse, especially in the context of national security. Drawing inspiration from the United Nations (UN) Atomic Nuclear Agency, which monitors countries' activities related to nuclear technology, the author proposed the establishment of a similar global entity to oversee the development and deployment of AI technology by nations.

In this paper, the research methodology revolves around conducting a thorough literature review on AI-related publications. The objective is to present a comprehensive analysis outlining the necessity for an international organization dedicated to monitoring AI activities. The paper explores several factors, strategies, and challenges associated with monitoring AI, leading to practical recommendations and suggesting potential areas for future research.

THE NEED FOR AI ACTIVITIES MONITORING ORGANIZATION

One of the primary reasons for establishing an organization to monitor AI activities in countries is the *potential misuse of AI for malicious purposes*. AI technology belongs to the category of dual-use products that may be used for a wide range of civilian needs. Dual-use AI technology capability also makes it applicable for

.

defense needs and consequently may pose security risks. AI systems can be programmed to undertake various tasks, from cyberattacks in connected networks to autonomous weapons, and the potential for misuse is enormous. The development of AI technology for military purposes is causing concern in the international community. Countries are investing heavily in developing AI-enabled weapons systems, and the lack of international regulation could lead to an arms race that could destabilize the global security order (Asaro, 2012; Gambrell, 2023).

Establishing an AI monitoring organization would help mitigate the risks associated with the *misuse of AI technology*. Such an organization could monitor the activities of countries and identify potential security threats arising from the development and deployment of AI systems. It could also provide guidance on the ethical and responsible use of AI technology and promote the development of international norms and standards for using AI in the military (Lopez, 2020).

Furthermore, an AI monitoring organization could promote *global cooperation in developing and deploying AI technology*. The development of AI is a complex and multifaceted process that requires expertise from various fields, including computer science, engineering, and mathematics. Establishing a platform for international cooperation could help to share knowledge and resources and promote innovation in the field. It could also help to bridge the gap between developed and developing countries and ensure that all countries have access to the benefits of AI technology.

Finally, an AI monitoring organization could promote *transparency and accountability in developing and deploying AI technology*. The lack of transparency in developing AI systems has raised concerns about the potential for bias and discrimination. An AI monitoring organization could promote transparency by requiring countries to disclose their AI development and deployment activities, including the algorithms and data used. It could also ensure that countries are accountable for how AI technology is used and that there are appropriate mechanisms for oversight and regulation (Bostrom &Yudkowky, 2014).

Regulation of the use of AI technology can be implemented at the national and international levels. To some extent, the principles and Governance structures of AI monitoring are similar to other monitored dual-use technologies.

Governance structures can include increased oversight at multiple levels to reduce risks posed by dual-use technologies in nuclear, biological, and information fields (Harris, 2016). These *Governance structures monitor technology transfer across defense and civil systems*. Such structures exist at the national and international levels. At the national level, many countries establish national requirements for export licensing of dual-use applications and products to protect national security interests and promote foreign policy objectives. For example, in the USA, the Bureau of Industry and Security (BIS) of the Department of Commerce monitors export control, which deals with national security and high technology issues.

ISSN: 1941-6679-On-line Copy

.

FACTORS FOR MONITORING POTENTIAL THREATS FROM AI

Just like with cybersecurity, there will be a need to efficiently discover unusual AI behavior to protect people, businesses, and organizations. Such unusual behavior may result from a design flaw or the malicious intent of the developer. Some principles used in cybersecurity defenses (e.g., NIST Cybersecurity Framework) can be adopted to develop AI-security capability.

AI Security Risk Assessment from Microsoft Security (Pearce et al., 2021) is one of the first attempts to profile a comprehensive perspective on AI system security, outline threats to critical AI assets, and provide guidance to conduct AI security assessments. The methodology of this assessment, based on survey business practices at business and government organizations, suggests three control ratings – *severity, likelihood*, and *impact*. These control ratings can help identify critical controls and propose guidelines for Machine Learning security assessment, which consists of three phases – data collection, processing, and model training.

REAL-TIME FACTOR

More recently, Gartner's Top Strategic Technology Trends (Gartner, 2023) proposed AI-driven security methodology (TRISM – Trust, Risk, Security Management) and use cases followed by critical actions for implementing AI security management. Explainability/Model Monitoring is one of the fundamental components of this methodology. The TriSM supports AI model governance, trustworthiness, fairness, reliability, robustness, efficacy, and privacy. Also, it cites the *use of AI for monitoring human behavior*, called applied observability. One example (from Tesla) is linking a proposed insurance amount to "observable" real-time driving behavior. The link of AI to real-time behavior is a critical component of Adaptive AI – the new type of AI that will further accelerate AI development and continuously keep AI aligned to enterprise goals in real-time (Chollet, 2019).

We must distinguish *static* and *dynamic* (time-dependent) data in monitoring AI activities. *AI may not always act the same way, and the real-time factor is critical* for identifying and correlating the events that point to risks. Therefore, data collected on observability in real-time, after being filtered, de-identified (if required by law), and summarized, will be very constructive for AI Activities Monitoring Organization.

HIERARCHY OF AI CONTROLS

An AI application can be instrumental in extracting actionable data from a large integrated pool of raw data from distributed networks of multiple interdependent

ISSN: 1941-6679-On-line Copy

.

sources. Monitoring such AI applications may present a complex problem – it will require factoring distributed network (data sources) complexity in the monitoring algorithm. Reducing the complexity of distributed data networks and data complexity for combating AI threats may not be feasible. Consequently, increasing complexity would lead to increased uncertainty (and inaccuracy) in AI solutions. An adversary can exploit such uncertainty. One solution to this problem is to have a controlling AI monitor the AI at a lower level, progressively leading to a *hierarchy of AI controls where the upper-level AI monitors one or more lower-level AIs*. Also, such a layered model of AI can be adopted for a framework for AI Governance, where a specific set of control criteria will define each layer. (Gasser and Almeida, 2017).

STRATEGIES FOR REGULATIONS

With the rapid proliferation of AI technologies (and their risks) in banking and public spheres, new AI regulations (national and international) for data protection are supported by existing laws (e.g., GDRP – General Data Protection Regulation) and technologies.

New regulations are needed to address online harms (risks) from an overwhelming reliance on AI solutions in combating online problems. Thus, in the USA, the *Federal Trade Commission (FTC), tasked to develop recommendations for the new AI regulation*, identified the following problematic areas of risk: *inherent design flaws, inaccuracy, bias, discrimination*, and *commercial surveillance creep*. (FTC, 2022) Complex issues surrounding these problems include the following: differences in business models used to combat online harms and lack of business policies, lack of access to proprietary AI tools, frequent inaccuracy of datasets characterized by content often too complex or subjective to determine whether harm can occur, dependency on context, and others. In its analysis, FTC points out that sometimes "increased accuracy could itself lead to other harms, such as enabling increasingly invasive forms of surveillance. The FTCs recommendations concluded that government and companies should "exercise great caution in either mandating the use of or over-relying on [AI] tools even for the important purpose of reducing harm" (FTC, 2022).

As a precursor to implementing the anticipated regulations, many leading hi-tech corporations are developing or implementing formal AI policies to protect safety and privacy (Candelon, Carlo, Bondt, and Evgeniou, 2021). For example, Facebook uses AI tools to fight fraud, though it does not stop scams from continually rising. Google uses AI to detect online fraud in searches and Gmail. AI helps detect fake online reviews (Facebook, Best Buy), counterfeit products (eBay, Etsy, Facebook, and Alibaba), and tax scams' (FTC, 2022)

.

AI monitoring, in some respect, is analogous to *cybersecurity monitoring*. Requirements for monitoring AI can be added to the existing and future AI Framework. Several currently available AI frameworks are well described in public sources. For example, TensorFlow supports deep learning accessible from desktop and mobile devices. Microsoft CNTK supports text, message, and voice remodeling of data on major massive datasets. Caffe is a deep-learning network with powerful image-processing capabilities (Kothari, 2023).

In June 2023, the European Union achieved a significant milestone by taking decisive action towards establishing pioneering regulations for the utilization of artificial intelligence by companies. This historic move by Brussels aims to create a framework that can serve as a model for global standards in the application of AI technology, spanning a wide range of fields including chatbots like OpenAI's ChatGPT, surgical procedures, and fraud detection in the banking sector (Ziady, 2023).

STRATEGIES FOR REGULATION OF DUAL-USE TECHNOLOGIES AT INTERNATIONAL ORGANIZATIONS

At the international level, the cross-border transfer of dual-use technologies is regulated based on bilateral agreements between countries or via international organizations. The United Nations, the leading global organization regulating global technologies, developed the *Strategy on New Technologies*.

Strategy on New Technologies at the UN, announced by the Secretary-General (UN Secretary-General, 2018; UN Office of Counterterrorism, 2023), states the objective to "support the use of these technologies to accelerate the achievement of the 2030 Sustainable Development Agenda." It also intends "to facilitate their alignment with the values enshrined in the UN Charter, the Universal Declaration of Human Rights, and the norms and standards of International Laws."

The Strategy emphasizes, among others, the role of artificial intelligence in "realizing the benefits of [new technologies] and helping contain the risks posed by new technologies." Specifically, it strives to uphold the values and obligations of the UN Charter and the Universal Declaration of Human Rights. It also provides a platform for governments, businesses, and civil society across generations to make collective choices about new technologies. These and other principles summarized in the Strategy are designed to strengthen the UN's capacity to engage with new technologies, increase UN outreach and engagement, promote cooperation frameworks, and support Member States. The task of monitoring AI at the UN level can be aligned with the existing roles of one of the Autonomous Organizations linked to the UN through special agreements.

The European Commission proposed its own framework on AI regulation with four objectives: safety as laws on fundamental rights concern, legal certainty for

.

investments and innovations in AI, effective enforcement of existing laws, and a single market for trustworthy AI applications, which would prevent market segmentation. The proposal envisions "a *single future-proof definition of AI*" and categorizes the use of AI by the levels of severity of a risk. (European Commission, 2021).

MONITORING OF AI USES

While AI regulations will *establish compliance rules*, AI monitoring will *test that compliance*. One of the first projects addressing AI monitoring capabilities is the European Red-Alert system (Red-Alert, 2023; Red-Alert, 2020). The system is proposed to assist European Union Law Enforcement Agencies (LEAs) in using AI, Natural Language Processing (NLP), Social Network Analysis (SNA), and Complex Event Processing (CEP) for analyzing social media intelligence in a wide range of social media channels. The Red-Alert project sets objectives to improve the results of NLP, SNA, and CEP by using AI capabilities and integrating them all for real-time collaboration.

On a more elementary level, YouTube (YouTube, 2022) and TikTok built automated capabilities to detect (and remove) violent and graphic content. There are other examples of using AI in real-time for monitoring online data. For example, Graphika, the social media company, monitors inappropriate images.

CHALLENGES FOR AI ACTIVITIES MONITORING AT INTERNATIONAL ORGANIZATION

Establishing an international organization to monitor the AI activities of countries poses several challenges. One of the main challenges is the need for more consensus on developing and deploying AI technology. Countries have different priorities and interests regarding AI development, and some countries may be reluctant to cede control over their AI activities to an international organization. Additionally, there needs to be a more collective understanding of what constitutes responsible AI development, and different countries may have different standards and values when it comes to AI development (UNESCO, 2023).

Another challenge is the difficulty of monitoring AI activities, as *AI algorithms and technologies are complex and rapidly evolving*. It may be challenging to keep up with the development of new AI technologies and ensure that they are being used responsibly. Furthermore, monitoring AI activities may require significant resources and expertise, and it may be challenging to build a network of experts and laboratories to support the organization's work (Dwivedi et al., 2021).

There is also the issue of *funding*, as establishing and operating an international organization to monitor AI activities would require significant financial resources.

ISSN: 1941-6679-On-line Copy

.

Securing funding from member countries may be challenging, especially given the current geopolitical landscape.

Finally, there may be concerns about the *potential misuse of the organization's powers*. Some countries may be hesitant to give an international organization the power to regulate their AI activities, and there may be concerns that the organization may be used to advance the interests of some countries over others.

PRACTICAL RECOMMENDATIONS

Despite the challenges, there are several potential solutions to establish an international organization to monitor the AI activities of countries. One potential solution is to focus on building consensus around responsible AI development. It could involve establishing *global standards* for AI development, promoting *transparency* in AI development, and encouraging *collaboration* between countries. Building consensus around responsible AI development may make establishing an international organization to monitor AI activities more feasible.

Another potential solution is to establish a *phased approach to establishing the organization*. It could involve starting with a smaller group of countries willing to participate in the organization and gradually expanding membership over time.

This approach would allow the organization to build momentum and demonstrate its effectiveness before expanding its reach.

To address concerns around funding, the organization could explore *alternative funding* models, such as public-private partnerships, to secure additional financial resources. The organization could also explore ways to leverage existing resources, such as partnering with existing research centers and experts in AI. Finally, to address concerns about the potential misuse of powers, the organization could establish a transparent and accountable governance structure. It could involve establishing an *oversight committee* to ensure that the organization operates within its mandate and promotes all member countries' interests.

Protecting the UN from AI-based vulnerabilities could be like protection from cyber-based vulnerabilities. The Office of Counterterrorism at the United Nations (UNOCT) has several initiatives in the modern technologies field. It deals with the misuse of information and communications technologies and multi-stakeholder cooperation (among Member States) in protecting against the threat of cyber-attacks. *Linking AI and cybersecurity in organizational structures* will help expedite and streamline the monitoring of inappropriate use of AI. One significant difference between these two is that AI can be a powerful tool to counter cybersecurity terrorism online (Sarker et al., 2021).

ISSN: 1941-6679-On-line Copy

.

RECOMMENDATIONS FOR FURTHER RESEARCH

Based on the paper's analysis of the necessity for an AI monitoring organization and strategies for regulating AI research and development activities, below are some recommendations for future research:

Governance and Accountability: Research can focus on developing frameworks and models for the governance and accountability of AI monitoring organizations. This includes exploring the legal and ethical implications of granting monitoring powers to an international organization and defining mechanisms for transparency, oversight, and accountability (Ziady, 2023).

International Cooperation: Further research can delve into the potential models of international cooperation in AI development and deployment. This includes investigating mechanisms for knowledge sharing, resource allocation, and collaboration between countries to ensure equitable access to AI benefits and minimize the risk of an AI arms race (Zhang et al., 2022).

Ethical and Responsible AI Development: Research can explore the development of ethical guidelines and standards for AI technology, particularly in the context of military applications. This includes addressing issues such as bias, discrimination, social media, privacy, and transparency in AI algorithms and systems developed for defense purposes (Axente, M. & Golbin, I., 2021).

AI Security and Risk Assessment: Future research can focus on developing robust methodologies for assessing the security risks associated with AI systems.
This includes identifying potential vulnerabilities, designing effective control measures, and developing frameworks for real-time monitoring and detection of unusual AI behavior (Dempsey, 2023).

Funding Models and Sustainability: Research can explore alternative funding models for AI monitoring organizations, such as public-private partnerships, to ensure their financial sustainability. Additionally, investigating strategies for leveraging existing resources and expertise in AI research centers can help optimize the organization's operations.

Consensus Building and Standards: Further research can investigate strategies for building consensus and establishing global standards for responsible AI development. This includes exploring approaches to harmonize diverse national interests, values, and standards related to AI technology (Ziady, 2023).

ISSN: 1941-6679-On-line Copy

.

The Role of Existing International Organizations: Research can assess the potential roles and responsibilities of existing international organizations, such as the United Nations and its specialized agencies, in monitoring AI activities. This includes examining the feasibility of integrating AI monitoring within these organizations' existing frameworks and mandates.

AI Monitoring Technologies and Tools: Future research can focus on the development of advanced technologies and tools specifically designed for AI monitoring. Just as we employ police officers to oversee human activities, it is imperative to have AI officers who can monitor and regulate the actions of other AIs. This includes exploring techniques for real-time data collection, analysis, and modeling to detect and respond to potential risks and threats arising from AI activities.

Authenticity of AI-generated content: Additional research can propose normative methods to assess the maturity of Generative AIs (as opposed to predictive AIs) in creating content that meets the quality objectives (keeping it free from misleading, false, or low-quality information) and authenticity requirements to prevent spreading of misinformation (such as using watermarks implanted into AI-generated content). (Shreyansh, 2023; White, 2023).

Trustworthiness of AI systems: A more encompassing research can investigate how the metrics and tools for AI systems align with Trustworthy AI principles (OECD, 2023; OECD, 2021; Saif & Ammanath, 2020). Fundamental principles of Trustworthy AI may include robustness, generalization, explainability, transparency, reproducibility, fairness, privacy protection, value alignment, and accountability. (Li et al., 2021)

These recommendations provide a starting point for future research in AI monitoring and regulation. They address the challenges highlighted in the paper and aim to foster the responsible and accountable development and use of AI technology at the national and international levels.

CONCLUSION

The deployment of AI technology has the potential to bring significant benefits to society, but it also poses many risks. To ensure the responsible use of AI technology, the establishment of an international organization to monitor the AI activities of countries is necessary. Such an organization would promote responsible AI development, establish global standards for AI development, and monitor the use of AI for harmful purposes. While establishing such an organization

ISSN: 1941-6679-On-line Copy

.

poses several challenges, there are potential solutions that could be explored to overcome these challenges. The establishment of an international organization to monitor AI activities is necessary to ensure the safe and responsible deployment of AI technology globally.

One of the major limitations of this research is the complexity and rapid evolution of AI technologies. AI algorithms and technologies are complex and constantly progressing, making it essential to continuously stay updated with the latest advancements and comprehend the implications associated with these technologies to effectively monitor AI activities.

REFERENCES

Asaro, P. (2012). On banning autonomous weapon systems: human rights, automation, and the dehumanization of lethal decision-making. Cambridge University Press: online published 24 June 2013. International Review of the Red Cross, Volume 94, Issue 886: New Technologies and Warfare, June 2012, pp. 687 – 709. https://doi.org/10.1017/S1816383112000768

Axente, M. & Golbin, I. (2021). Ethical AI: 10 principles the world (mostly) agrees on — and what to do about them. August 12, 2021. https://www.pwc.com/us/en/tech-effect/ai-analytics/how-to-make-ai-ethical.html.

Bostrom, N. & Yudkowsky, E. (2014). The ethics of artificial intelligence. In *The Cambridge Handbook of Artificial Intelligence* (pp. 316-334), June 2014. Cambridge University Press. https://www.fhi.ox.ac.uk/publications/bostrom-n-yudkowsky-e-2014-the-ethics-of-artificial-intelligence-the-cambridge-handbook-of-artificial-intelligence-316-334/

Candelon, D. C. et al (2021). AI Regulation Is Coming. *Harvard Business Review*. https://hbr.org/2021/09/ai-regulation-is-coming.

Chollet, F. (2019). On the Measure of Intelligence. Google, Inc. https://arxiv.org/pdf/1911.01547.pdf

Dempsey, J. (2023). Addressing the Security Risks of AI, https://www.lawfareblog.com/addressing-security-risks-ai.

Dwivedi, Y. et al. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for

.

research, practice and policy. Elsevier. *International Journal of Information Management*, Volume 57, April 2021, 101994. https://www.sciencedirect.com/science/article/abs/pii/S026840121930917X?via%3Dihub

European Commission (2021). Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). European Commission. *Proposal for a Regulation of The European Parliament and of The Council*. Document 52021PC0206. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206.

FTC (2022). Combating Online Harms Through Innovation: A Report to Congress, June 16, 2022. Federal Trade Commission. https://www.ftc.gov/system/files/ftc_gov/pdf/Combatting%20Online%20Harms%20Through%20Innovation%3B%20Federal%20Trade%20Commission%20Report%20to%20Congress.pdf.

Gambrell, J. (2023). OpenAI CEO suggests international agency like UN's nuclear watchdog could oversee AI. June 6, 2023. www.candorium.com/news/20230606151027599/openai-ceo-suggests-international-agency-like-uns-nuclear-watchdog-could-oversee-ai.

Gartner (2023). Top Strategic Technology Trends 2023. Gartner. https://enterprise.press/wp-content/uploads/2022/10/Gartner.pdf

Gasser, U. and Almeida, V. (2017). A Layered Model for AI Governance. Harvard University. https://dash.harvard.edu/bitstream/handle/1/34390353/w6gov-18-LATEX.pdf?sequence=1.

Harris, E. et al (2016). Governance of Dual-Use Technologies: Theory and Practice. Project: Global Nuclear Future. Research Paper, April 2016. American Academy of Arts & Sciences. https://www.amacad.org/publication/governance-dual-use-technologies-theory-and-practice.

Kothari, A. (2023). Top 11 Frameworks in the AI World. GeekFlare. https://geekflare.com/ai-frameworks/.

Li, B. et al. (2021). Trustworthy AI: From Principles to Practices. October 2021. *Association for Computing Machinery*. Vol. 1, No. 1, Article. https://www.researchgate.net/publication/355060788_Trustworthy_AI_From_Principles_to_Practices.

ISSN: 1941-6679-On-line Copy

.

Lopez, T. (2020). DOD Adopts 5 Principles of Artificial Intelligence Ethics, U.S. Department of Defense, Feb. 25, 2020. https://www.defense.gov/News/News-Stories/Article/Article/2094085/dod-adopts-5-principles-of-artificial-intelligence-ethics/.

OECD (2021). Tools for Trustworthy AI: A Framework to Compare Implementation Tools for Trustworthy AI Systems. *OECD Digital Economy Papers*, June 2021 No. 312. OECD Publishing.

OECD (2023). Catalogue of Tools & Metrics for Trustworthy AI, AI Policy Observatory. OECD.org. https://oecd.ai/en/catalogue/metrics

Pearce, W. et al. (2021). AI Security Risk Assessment: Best practices and guidance to secure AI systems. Microsoft Security. Dec 2021. https://github.com/Azure/AI-Security-Risk-Assessment/blob/main/AI_Risk_Assessment_v4.1.4.pdf

Red-Alert (2020). Real-time Early Detection and Alert System for Online Terrorist Content based on Natural Language Processing, Social Network Analysis, Artificial Intelligence and Complex Event Processing. *Research & Innovation for Secure Societies*. https://www.enisa.europa.eu/events/artificial-intelligence-an-opportunity-for-the-eu-cyber-crisis-management/workshop-presentations/20190604-bcu-red-alert-dissemination.pdf

Red-Alert (2023). Real-Time Early Detection and Alert System. *Red Alert Project*. https://redalertproject.eu/.

Saif, I. and Ammanath, B. (2020) The Trustworthy AI framework. The Deloitte AI Institute. MIT Technology Review. Artificial Intelligence. https://www.technologyreview.com/2020/03/25/950291/trustworthy-ai-is-a-framework-to-help-manage-unique-risk/

Sarker, I. et al. (2021). AI-Driven Cybersecurity: An Overview, Security Intelligence Modeling and Research Directions. https://link.springer.com/article/10.1007/s42979-021-00557-0#citeas.

Seger, E. et al. (2023). Centre for the Governance of AI Oxford, UK. "Democratising AI": Multiple Meanings, Goals, and Methods, https://arxiv.org/ftp/arxiv/papers/2303/2303.12642.pdf.

ISSN: 1941-6679-On-line Copy

.

Shreyansh, K. (April 2023). Generative AI and digital content authenticity. https://bootcamp.uxdesign.cc/generative-ai-and-digital-content-authenticity-fc44c2c4580.

UN Office of Counterterrorism (2023). United Nations Office of Counter-Terrorism. Cybersecurity and New Technologies. https://www.un.org/counterterrorism/cybersecurity.

UN Secretary-General (2018). Strategy on New Technologies. https://www.un.org/en/newtechnologies/

UNESCO (2023). International cooperation is key to inclusive AI. https://www.unesco.org/en/articles/international-cooperation-key-inclusive-ai.

White, S. (March 2023). Watermarking ChatGPT, DALL-E and other generative AIs could help protect against fraud and misinformation, *The Conversation*. https://theconversation.com/watermarking-chatgpt-dall-e-and-other-generative-ais-could-help-protect-against-fraud-and-misinformation-202293.

YouTube (2022). YouTube Community Guidelines enforcement. https://transparencyreport.google.com/youtube-policy/removals.

Zhang, D. et al (2022). Enhancing International Cooperation in AI Research: The Case for a Multilateral AI Research Institute, White Paper, May 2022, Stanford University, Human-Centered Artificial Intelligence. https://hai.stanford.edu/white-paper-enhancing-international-cooperation-ai-research-case-multilateral-ai-research-institute.

Ziady, H. (2023). Europe is leading the race to regulate AI. Here's what you need to know. https://www.cnn.com/2023/06/15/tech/ai-act-europe-key-takeaways/index.html.

ISSN: 1941-6679-On-line Copy

.