Peace Research Center Prague

Policy Brief 023

March 2024 www.prcprague.cz

The Prospects and Limitations of the Nuclear-AI Analogy

Karim Kamel

In May 2023, the United States Senate Judiciary Subcommittee on Privacy, Technology and the Law held a pivotal hearing on Artificial Intelligence (AI). The hearing included a testimony from Sam Altman, CEO of OpenAI–the creators of ChatGPT. During the meeting, referencing an earlier remark by Senator Lindsey Graham, Senator Peter Welch said "you don't build a nuclear reactor without getting a license; you don't build an AI system without getting a license that gets tested independently." To this Sam Altman promptly responded, "I think it's a great analogy."

There are two ways experts and policy stakeholders have articulated the nuclear-AI analogy. First is humanity unveiling a magnificent force and facing a defining moment. Just like nuclear technology, the unearthing of AI is disruptive, and it could be used for good or for ill. The second analogy relates to practical lessons for AI governance from the nuclear experience. Nuclear governance includes regulations, agencies, verification measures and partnerships, which could also be applicable to AI.

The analogy has merits. However, differences in context, technological composition, and the potential benefits and risks, will require a pluralistic approach to AI beyond the model that nuclear technology offers.

On the unleashing of a force; on the defining moment

It is true that both the nuclear and the AI moments are groundbreaking. By splitting the atom, we unleashed new frontiers of power that could fuel human development or destroy cities. The main technological breakthrough of AI–Large Language Models–is just as powerful. Generative AI holds the promise to identify research gaps, propose studies and experiments, as well as conduct them. This automation of research could be particularly groundbreaking in the medical space, but that same generative AI could provide instructions on how to synthesize a chemical or biological weapon.[1]

[1] Urbina, F., Lentzos, F., Invernizzi, C. et al. Dual use of artificial-intelligence-powered drug discovery. Nat Mach Intell 4, 189–191 (2022). https://doi.org/10.1038/s42256-022-00465-9

About the author

Karim Kamel is a Ph.D. student of Area Studies at the Faculty of Social Sciences, Charles University. In his research, he focuses on the the employment of foresight methodologies to analyze the impact of disruptive technologies on international peace and security.

While unearthing both technologies did present a defining moment in human history, the context is quite different. Splitting the atom came about amidst a world war. It was born out of a military program, and its main purpose was to manufacture a device for killing humans. AI on the other hand, has been born during an era of relative peace. AI was also developed by innovative non-governmental labs. In fact, OpenAI is a nonprofit, and its mission is "to ensure that artificial general intelligence benefits all of humanity." Thus, AI does not have the weaponization legacy of nuclear technology.

In this context, our approach to AI needs to avoid the pitfalls of over-securitization and a myopic focus on competition. We have already seen AI policies being crafted emphasizing "competition" with China. This included the U.S. "Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence" as well as the Department of Commerce caps on the export of advanced semiconductors to China. These policies have been described, somewhat favorably, as "choking off China's access to future AI." [2] These policies, while intended for national security, could inadvertently stifle the collaborative and pluralistic development necessary for global AI advancement.

AI is a true challenge to humanity, but it is unnecessary and unwise to lock ourselves into "a new arms race," which the UN High-Level Advisory Body (HLAB) on AI warns about in its latest report. [3] Even seemingly benign sentiments such as "America's got to continue to lead," as expressed by Sam Altman during the U.S. Senate hearing, should be downplayed.

The approach to AI needs to be global and pluralistic. As articulated by the UN HLAB AI, the vision for governance should be "a 'distributed-CERN' for AI, networked across diverse states and regions."

On governing AI like we govern the atom

The second part of the analogy relates to nuclear governance being a model for AI. The global response to nuclear technology, particularly the formation of the International Atomic Energy Agency (IAEA), offers valuable lessons for AI governance. However, applying IAEA safeguarding principles to AI is not a straightforward prospect. Unlike nuclear material, which can be quantified and monitored, AI's components—data, algorithms, and computational power—are more elusive, with computational power being the most tangible for governance. Mauricio Baker recently published a study [4] on "Nuclear Arms Control Verification and Lessons for AI Treaties." The study identifies "training machine learning models with industrial-scale, specialized computer chips" as the technology to control. Compute being the physical product in AI makes it governable based on controlling access to it measured in FLOPS. This hardware-enabled approach [5] is useful but not sufficient.

While controlling the production and distribution of advanced computing chips is feasible, the rapid advancements in AI technology make setting a fixed standard for computation power impractical.

[2] Allen, G.C. Choking Off China's Access to the Future of AI: New U.S. Export Controls on AI and Semiconductors Mark a Transformation of U.S. Technology Competition with China. Center for
[3] Strategic and International Studies (2022).
United Nations High-Level Advisory Body on AI. Interim Report: Governing AI for Humanity. December 2023. www.un.org/en/ai-advisory-body.
[4] Baker, M. Nuclear Arms Control Verification and Lessons for AI Treaties. arXiv:2304.04123v1 [cs.CY] (2023).
[5] TechTarget. Floating-point operations per second (FLOPS). Retrieved from https://www.techtarget.com/whatis/definition/FLOPS-floating-point-operations-per-second#:~:text=Floating%2Dpoint%20operations%20per%20second%20(FLOPS)%20is%20a%20measure,can%20perform%20within%20a%20second.

Nuclear safeguards orient their work around significant quantities of fissile material, which are 8 kg PU, 8 kg U233, and 25 kg U235. A significant quantity in compute will always be a moving target, and will depend on improvements in algorithms. For the same reasons, export control, as another hardware-enabled governance tool, may soon become outdated. Forming exclusive clubs akin to the Nuclear Suppliers Group cannot be done in isolation from the global community. Thus, these approaches must be complemented by a broader political agreement, akin to the nuclear nonproliferation regime's "grand bargain."

Image generated by AI, courtesy of OpenAI's DALL-E

Grand Bargain not Grand Strategy

The cornerstone instrument of the nuclear nonproliferation regime, the Treaty on the Nonproliferation of Nuclear Weapons (NPT), is based on what diplomats refer to as the "grand bargain." Nuclear Weapon States committed to sharing the peaceful benefits of the technology while not spreading the military applications to new actors. Non-Nuclear Weapon States, on the other hand, promise not to pursue nuclear military capabilities. What could a grand bargain for AI look like?

Capping computing power and enforcing export control on sensitive components would be accompanied with transparency and a "sharing-the-benefits" roadmap. And on data, we need to ensure that equitable representative data is used to train AI models.

During the UN Security Council meeting on AI, the representative of Brazil rightly noted that in LLMs "the outcomes are crucially dependent on input… otherwise we run the risk that the aphorism 'trash in, trash out' becomes a self-fulfilling prophecy." To avoid the "trash in, trash out" scenario, we must be vigilant in curating[15] the data that feeds AI, ensuring it is representative, unbiased, and conducive to positive outcomes. The call for an inclusive international process is not just about regulation, but about cultivating a global discourse on AI that reflects diverse perspectives and values. It is thus crucial to be cognizant of how we populate the discourse, posit the questions, collect the data, and choose directions.

Global, plural, and human-aligned AI is already emerging through the work of a multitude of UN agencies. In 2022, 40 UN agencies reported 281 AI-related projects.[6] Most of these projects could be of great benefit to humanity, such as the UN's Department of Peace Operations, "Innovation Cell" for "exploring the use of AI for mediators and actors to hold real-time consultations with a large group of individuals in local dialects and languages." Humanity's biggest advantage is our ability[16] to cooperate. This is why the UN specifically collects data on multi-stakeholder collaborations on AI. Of the 281 projects, only 25% reported cooperation with academia or governments, and 20% reported cooperation with the private sector.

[6] International Telecommunication Union. United Nations Activities on Artificial Intelligence (AI). [online] Available at: https://www.itu.int/pub/S-GEN-UNACT (2022)

We need to prioritize multi-stakeholder engagements, especially among governments, academia, and the private sector. The private sector has been particularly absent from the Convention of Certain Conventional Weapons discussions on defining "Lethal Autonomous Weapon Systems." These Geneva discussions have been stagnant for years. We ought to take stock promptly of stakeholders' positions on the deployment of black-box lethal systems. These discussions should bring the 127 member states of the Convention together with stakeholders from academia and the private sector to get the process unstuck. A multi-stakeholder engagement ought to handle the representation and the mechanics of the discussions with intention. A truly multicultural discussion should be rooted in mediation,[7] discussion management,[8] and value elicitation processes.[9]

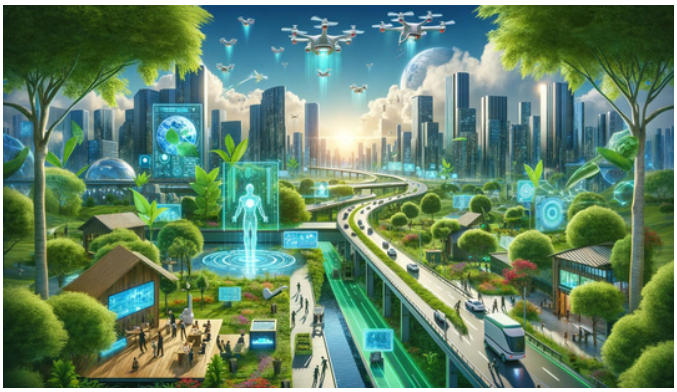

Image generated by AI, courtesy of OpenAI's DALL-E

At the same Security Council meeting on AI, the representative of Ecuador cited the Polish writer Stanislaw Lem, "the primary obligation of intelligence is to distrust itself." In the spirit of Stanislaw Lem's wisdom, we must approach AI with a healthy dose of skepticism and humility, recognizing both its immense potential and its inherent risks. The aim is not to hinder AI's progress but to steer it in a direction that uplifts humanity, fosters global collaboration, and addresses the pressing challenges of our time.

[7] RapidRuling. (n.d.). Mediation Techniques: Tools for Effective Conflict Resolution. Retrieved from https://rapidruling.com/blog/alternative-dispute-resolution-blog/mediation-alternative-dispute-resolution-blog/mediation-techniques-tools-for-effective-conflict-resolution/

[8] Cloke, Ken, Wendy Wood, and Scott Martin. 'Dialogue and Facilitation: Tools for Generative Conflict & Resilient Groups'. Mediators Beyond Borders International.

[9] Verdiesen, I., Dignum, V. Value elicitation on a scenario of autonomous weapon system deployment: a qualitative study based on the value deliberation process. AI Ethics