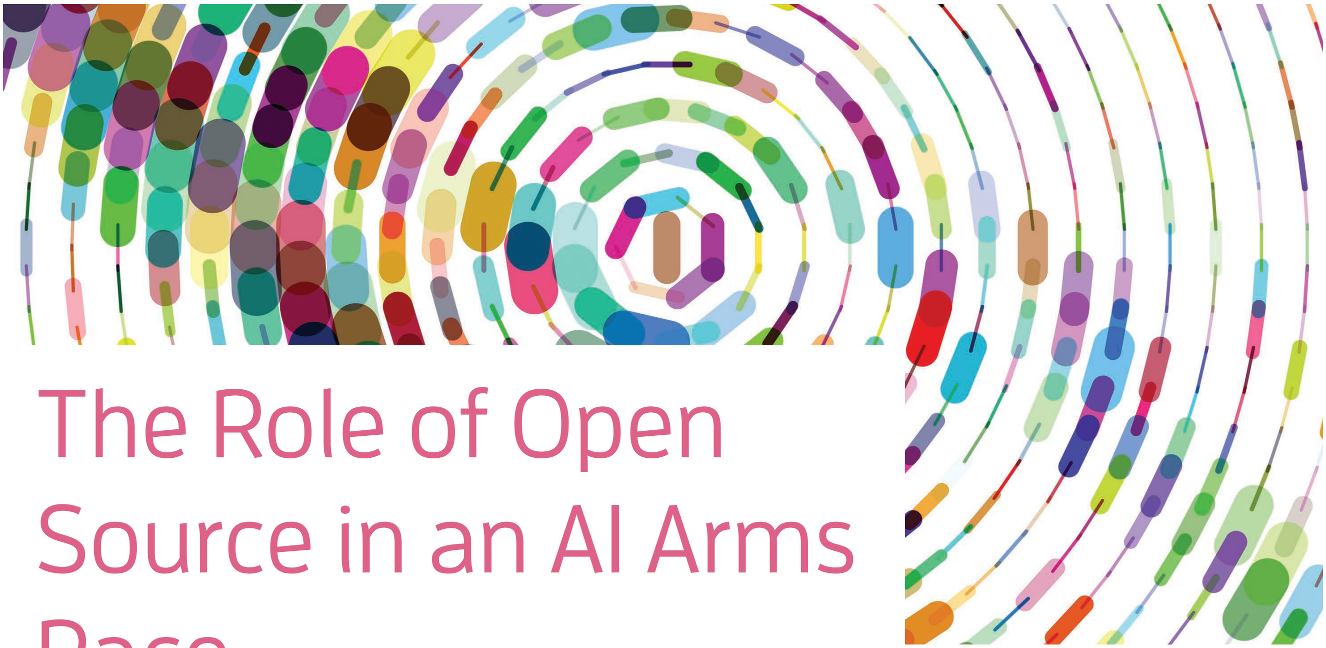
EDITOR DIRK RIEHLE
Friedrich Alexander-University of Erlangen Nürnberg;
dirk.riehle@fau.de



The Role of Open Source in an AI Arms Race

Christian Koch, Nuremberg Institute of Technology Georg Simon Ohm

Advances in artificial intelligence could trigger a new international arms race. This article discusses the role of open source software in this contest from a perspective of national defense.

In the 2023 movie "Oppenheimer" there is a pivotal scene.[1] The main character Robert Oppenheimer debates with Albert Einstein whether a nuclear explosion could ignite the atmosphere and destroy Earth. Einstein's character responds with a remarkable suggestion: If a mathematical analysis supports the hypothesis, American scientists should share their findings with Nazi Germany so that "neither side destroys the world." Whether this conversation really took place can be doubted. What is certain is that the possibility of an atmospheric ignition was discussed in the Manhattan Project.[2] Furthermore, leading physicists like Niels Bohr proposed a policy of openness toward nuclear science. Their core idea was to share knowledge about nuclear weapons with competitors like the Soviet Union to avoid a nuclear arms race.[2] Such an arms race later unfolded during the Cold War culminating in a doctrine known as "mutual assured destruction."[3]

Our generation could face a similar dilemma with artificial intelligence (AI). While many problems with nuclear weapons remain unsolved, AI is about to trigger another revolution in warfare.[a] According to Kissinger et al., the AI era "risks complicating the riddles of modern strategy further beyond human intention—or perhaps complete human comprehension."[3] In their book *The Coming Wave*, Suleyman and Bhaskar discuss the risks of AI for modern society. The authors conclude that "total openness to all experimentation and development is a straightforward recipe for catastrophe." In their view, open source "has been a boon to technological development" but is "not an appropriate philosophy for powerful AI models."[4] On the other hand, the authors emphasize that "there is no path to

[a] Many see an AI arms race already in full swing.

Digital Object Identifier 10.1109/MC.2025.3581000
Date of current version: 25 September 2025

FROM THE EDITOR

Welcome back to our column on open source! Would you like to contribute? If so, please send me an email with your article idea. In no time you might be joining Christian Koch, our author for this column. He reviews the role open source software in a world of AI and finds benefits and drawbacks, as to be expected. Happy hacking, and stay happy and healthy.—*Dirk Riehle*

technological safety without working with [one's] adversaries."[4]

This leads us to the question of what information we should share with adversaries. When is open source software (OSS) a useful instrument in an AI arms race and when is it counterproductive? This essay discusses the question from a perspective of national defense.

SECURITY AND SAFETY

At the heart of national defense is the protection of people from harm. The Basic Law of the Federal Republic of Germany enshrines this goal in its oath of office. Every Federal President, Chancellor, and Minister must swear to "protect [German people] from harm."[5] To fulfill this obligation, a government needs to assess the risks associated with its actions. The main hypothesis of this essay is that a risk-based approach should define the policy toward open source in an AI arms race.

When analyzing risk, it is helpful to distinguish between *security* and *safety*. Security is concerned with intended actions while safety deals with unintended consequences.[6] Let us consider some examples. Most military attacks are carried out by intention and are thus matters of national security. This includes both symmetric and asymmetric threats. On the other hand, the accidental outbreak of a biological warfare agent from an adversary's laboratory is unintended and thus a safety issue. National defense must take both risk types into account—intended and unintended.

Einstein's fictitious recommendation from the introduction makes it clear that there can be a tradeoff between security and safety. In this example, sharing information on how to destroy Earth would have increased America's safety by preventing Nazi Germany from causing a global disaster. However, the information could at the same time have weakened America's security. Sharing the knowledge on how to ignite the atmosphere would in effect have put a doomsday machine in the hands of a rogue government. Of course, the example represents a most extreme case. Yet the metaphor points to a fundamental dilemma we face with software in an AI arms race.

OPPORTUNITIES AND RISKS

When it comes to artificial intelligence, OSS is relevant in two ways.[b] First, we can use it to train and execute AI models. This includes general-purpose technologies like programming languages and development environments but also specialized tools with focus on machine learning. The second way is to provide the weights[c] of a trained AI model as open source, also called *open weight*.[7] One example for the latter is Meta's large language model Llama.[8] Users have full transparency about Llama's weights and can download them. An alternative are closed-source AI models, where the system's inner workings are kept secret.[d,e]

In the face of an AI arms race, open and closed-source software (CSS) entail specific opportunities and risks. According to the U.S. Department of Defense (DoD), open source "forms the bedrock of the software-defined world and is critical in delivering software faster."[9] The European Union describes OSS as a contribution to "research and innovation" and as a provider of "growth opportunities."[10] Alongside these benefits, there are also risks. These depend on the underlying asset. For example, an AI development environment poses other risks than the weights of a large language model. Typical concerns about OSS from a defense perspective are that it discloses innovation to adversaries and opens up the possibility for exploits.[9] When it comes to CSS, a potential vendor lock-in and a lack of transparency are main risks to consider.[8]

From the standpoint of national defense, an informed decision on whether a particular AI technology should be open source must be based on a risk assessment. Governments must weigh the upsides of OSS against its downsides for each case. Even missing out on the opportunities of open source poses a risk in itself. Whenever OSS reduces the expected harm to people, governments must utilize it. Conducting a risk assessment, however, is difficult.

When we seek guidance on the risks of AI systems, existing frameworks provide a starting point. One example of a risk-based regulation is the EU AI Act.[10] Adopted in 2024, the law sets out rules for nonmilitary AI applications in the European Union. It defines several risk classes with stricter regulations for high-risk systems in areas like health care, law enforcement, and critical infrastructure. Furthermore, it considers some practices as "unacceptable risk" and prohibits them. Examples of the latter are biometric categorization and social scoring. OSS is, in general, viewed positively in the EU AI Act. As a consequence, the Act allows exemptions for OSS, although not for high-risk systems and unacceptable risk.

Being a relatively new regulation, there is an ongoing debate about how

[b] Of course, other factors such as access to hardware and data are also crucial in an AI arms race from a perspective of national defense. But these are outside the scope of this essay.

[c] In a neural network, a decision is formed by multiplying inputs with "weights" that determine how strongly each input influences the final result.

[d] Status at the time of writing (July 2025).

[e] In accordance with their license agreements, not all models may be used in military applications.

to implement the EU AI Act correctly. We can expect this initial phase to continue for some time with organizations gradually finding their way to adequately implement the law. Meanwhile, we can learn from other sectors outside of artificial intelligence. One example of a global safety standard from another domain is provided by the Biosafety Levels.[4] This framework defines protective measures for biomedical laboratories depending on their risk level. Another field with a rich tradition in risk management is finance. Here we find principles that help us to mitigate the downsides of open source in an AI arms race.

MITIGATION STRATEGIES

In a contested environment, the impact of open source on AI is complex. In some cases, OSS will support national defense, while in others it is counterproductive. For many AI safety tools, for example, it is probably a good idea to share them with adversaries. On the other hand, open source tools for offensive purposes should likely be prohibited. But Einstein's story from the introduction warns us that the decision can require a tradeoff. This applies in particular to multiuse technologies. Manhattan Project member John von Neumann concluded in 1955 with nuclear weapons in mind that any "attempt to find automatically safe[f] channels for the present explosive variety of progress must lead to frustration. The only safety possible is relative, and it lies in an intelligent exercise of day-to-day judgment."[11] One core assumption of this article is that the same is true for open source in an AI arms race.

This essay neither advocates complete openness nor total secrecy toward AI. Instead, it proposes a risk-based approach. From a perspective of national defense, understanding the downsides of a particular AI tool is the only way to make an informed decision on whether OSS is a good idea. Governments must try to seize the opportunities of OSS while protecting themselves from its drawbacks. Risk assessments offer a structured path to identify potential threats, evaluate their probability and impact,[g] and define mitigation strategies.

Typical concerns about OSS from a defense perspective are that it discloses innovation to adversaries and opens up the possibility for exploits.

One principle governments can utilize to mitigate risks in an AI arms race is *diversification*. From the financial sector we know that this can be an effective method to counter individual downsides. Portfolio selection centers on the idea of combining assets in a way that balances out their specific risks. We can apply the same principle to AI systems. Diversification between open and closed-source modules can help to improve the resilience of system architectures and whole ecosystems.

One related concept is *redundancy*. Although redundancy and diversification are similar techniques, they are not identical. While diversification aims to spread risks across different modules of a system, redundancy is based on a duplication of components. The main goal of redundancy is to avoid a system breakdown if a particular element fails. Diversification on the other hand, aims to prevent correlated failures by using varied components. Both principles can form the basis of a risk mitigation strategy when it comes to sourcing in an AI arms race.[h]

Implementing a strategy of diversification and redundancy requires system architectures capable of integrating modules from multiple sources. One example of such an initiative is the Modular Open Systems Approach (MOSA) of the U.S. Department of Defense. The core idea of MOSA regarding OSS is outlined in a 2022 memorandum to the senior leadership of the Pentagon.[9] According to this letter, OSS accelerates development in the DoD but involves two key risks mentioned previously: the threat of exposing sensitive innovations to adversaries and the potential for exploits. MOSA's modular approach mitigates these downsides by requiring systems that are able to integrate open source while isolating critical components as separate closed-source modules. This combination of OSS and CSS allows protecting sensitive innovations while at the same time reducing the attack surface for potential threats.[i]

For a strategy of diversification and redundancy to work, modules combined must be independent. Otherwise, "single points of failure" arise in the architecture. To avoid this, it is necessary to understand the software supply chain of the modules used. One instrument to achieve this goal are software bills of materials (SBOMs). These provide an inventory of all components, libraries, and dependencies included in a specific software product.[14] With regard to artificial intelligence, specific DataBOMs or AIBOMs offer additional information about aspects like data sources, algorithms, and licenses. For future research, the development of tools and standards for this purpose is a promising area. As in the age of AI, transparency across the

[f] In his essay, von Neumann does not explicitly distinguish between security and safety.

[g] This article draws on the classic definition, where risk is the product of the probability and impact of a threat.

[h] In information security, the principles of diversification and redundancy find their realization among others in the concepts of compartmentalization and defense-in-depth.[12,13]

[i] Of course, CSS can also be known to adversaries. As von Neumann points out, success of secrecy is "never more—nor intended to be more—than transient."[11]

software supply will become a critical success factor for national defense.

Besides open source, there are many other aspects to discuss when it comes to an AI arms race. Among the most obvious are access to hardware, data, and personnel. But these are beyond the scope of this essay. One more holistic treatise on the subject is the Superintelligence Strategy by Hendrycks, et al.[7] Even in the area of open source, there are more issues to discuss than those covered previously. The main hypothesis of this essay is that there is no one-size-fits-all answer to the question of whether OSS is a useful instrument in an AI arms race. Decisions must be taken case by case based on a risk assessment. Only profound analysis will enable governments to benefit from the upsides of open source while protecting against its downsides.

Implementing a strategy of diversification and redundancy requires system architectures capable of integrating modules from multiple sources.

Frameworks like the EU AI Act help to identify potential threats and to find strategies of risk mitigation. One way to strengthen the resilience of AI applications is to strike a balance between open and closed source. Inner source can provide a compromise between these two poles. Implementing a strategy of diversification and redundancy requires modular system architectures and transparency across the software supply chain. SBOMs are one instrument for achieving this goal. Using these tools is essential for finding a risk-optimal mix of sources.

When it comes to national defense, the answer to the question of whether a particular AI tool should be open or closed source will often be ambiguous. Einstein's recommendation from the introduction illustrates this dilemma. In 1955, John von Neumann warned us that there is no general recipe for survival in a rapidly progressing world. He concluded that "we can specify only the human qualities required: patience, flexibility, intelligence." In 2025, we need exactly these qualities to prevail in an AI arms race. When it comes to OSS, there is no alternative to doing the hard work and examining the details of the technologies we aim to use.

REFERENCES

1. C. Nolan, "Oppenheimer," *Universal Pictures*, 2023. [Online]. Available: https://www.imdb.com/title/tt15398776/
2. K. Bird and M. J. Sherwin, *American Prometheus: The Triumph and Tragedy of J. Robert Oppenheimer*. London, U.K.: Atlantic Books, 2023.
3. H. A. Kissinger, E. Schmidt, and D. Huttenlocher, *The Age of AI: And Our Human Future*. New York, NY, USA: Little, Brown and Company, 2021.
4. M. Suleyman and M. Bashkar, *The Coming Wave. Technology, Power, and the 21st Century's Greatest Dilemma*. New York, NY, USA: Crown, 2023.
5. "Basic law for the Federal Republic of Germany." Federal Office of Justice, Dec. 19, 2022. Accessed: Dec. 28, 2024. [Online]. Available: https://www.gesetze-im-internet.de/englisch_gg/englisch_gg.html
6. B. Schneier, *Beyond Fear: Thinking Sensibly about Security in an Uncertain World*. New York, NY, USA: Copernicus Books, 2003.
7. D. Hendrycks, E. Schmidt, and A. Wang, "Superintelligence strategy: Expert version," vol. 2, no. 1, pp. 4–33, 2025, doi: 10.70777/si.v2i1.13961.
8. M. Zuckerberg, "Open source AI is the path forward," *Meta*, Jul. 23, 2024. [Online]. Available: https://about.fb.com/news/2024/07/open-source-ai-is-the-path-forward/
9. "Software development and open source software," U.S. Dept. Defense, Washington, DC, USA, Jan. 2022. [Online]. Available: https://dodcio.defense.gov/portals/0/documents/library/softwaredev-opensource.pdf
10. "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)," *Official J. Eur. Union Law*, vol. 202, pp. 1–145, Jul. 2024. [Online]. Available: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng
11. J. von Neumann, "Can we survive technology?" *Fortune*, 1955. [Online]. Available: https://sseh.uchicago.edu/doc/von_Neumann_1955.pdf
12. B. Schneier, *A Hacker's Mind: How the Powerful Bend Society's Rules, and How to Bend Them Back*. New York, NY, USA: Norton, 2023.
13. C. Koch, "The Prisoner's dilemma of open-source software security," *Computer*, vol. 57, no. 10, pp. 82–85, Oct. 2024, doi: 10.1109/MC.2024.3415868.
14. T. Stalnaker, N. Wintersgill, O. Chaparro, M. Di Penta, D. M. German, and D. Poshyvanyk, "BOMs away! Inside the minds of stakeholders: A comprehensive study of bills of materials for software systems," in *Proc. IEEE/ACM 46th Int. Conf. Softw. Eng. (ICSE)*, New York, NY, USA: ACM, 2024, pp. 1–13, doi: 10.1145/3597503.3623347.

CHRISTIAN KOCH is an enterprise lead architect at BWI GmbH and a lecturer at the Nuremberg Institute of Technology Georg Simon Ohm, 90489 Nuremberg, Germany. Contact him at christian.koch@th-nuernberg.de.