Reuters Legal News

Legal primer on open genAI models

By Ken D. Kumayama, Esq., and Pramode Chiruvolu, Esq., Skadden, Arps, Slate, Meagher & Flom LLP

AUGUST 15, 2024

With the release of ChatGPT, an era of generative artificial intelligence (genAI) began, and along with it, a spectrum of genAI models — from open to closed — has emerged. In this article, we discuss how the "openness" of various components of genAI models offers benefits and drawbacks with respect to transparency, control, innovation and usability, along with legal risks counsel should consider.

What makes a genAI model "open" remains subject to debate. For example, the April 2024 draft Open Source AI Definition (OSAID) from the Open Source Initiative suggests an open artificial intelligence system is one made available under terms granting the freedoms to use, study, modify and share the system for any purpose.

The "openness" of various components of genAI models offers benefits and drawbacks with respect to transparency, control, innovation and usability, along with legal risks counsel should consider.

The draft OSAID further specifies that access to enough information about the training data to allow a skilled person to recreate the system, the source code to train and run the system and the weights or parameters of the system are preconditions to exercise of those requisite freedoms.

Most practitioners have not adopted the OSAID and instead refer to open genAI systems more loosely — typically, open refers to models with weights published with the code to run the model, and "closed" or "proprietary" models are those where weights are unpublished. But use of such model weights and code for open genAI models may be subject to various restrictions, including restrictions on commercial use or modification, that are not in keeping with the freedoms listed in the draft OSAID.

Details regarding the data and training code of such open models may not be accessible, making reproduction of the model impracticable. Thus, referring to a genAI model as open or "open source" leaves out many of the details regarding the scope of freedoms and access.

Closed genAI models typically maintain not just their weights, but also the underlying architecture, source code and training data, as proprietary and confidential. Access to these models is usually provided through APIs (Application Programming Interfaces, which enable software components to communicate with each other). APIs allow external software to utilize the model's capabilities without disclosing the underlying source code or model weights.

These APIs usually run on remote backend servers optimized for performance and more powerful than hardware typically available to end users, allowing companies to deploy larger models, safeguard their intellectual property, charge for access, control use cases and implement safety checks. Examples include GPT-4***o*** from OpenAI and Gemini Pro from Google.

In contrast, open models are released based on a philosophy of transparency and collaboration. The source code used to train and run the model — and often the pre-trained model weights — are publicly accessible, fostering an environment where the community can inspect, improve and innovate.

While many open models are also available through APIs, these models can often be run on local, consumer-grade hardware, which allows for greater flexibility and customization and mitigates concerns regarding third-party access to sensitive data. Examples include OpenAI's Whisper for speech-to-text applications and Google's Gemma-2-27B.

Closed genAI models typically maintain not just their weights, but also the underlying architecture, source code and training data, as proprietary and confidential.

Proponents of open models argue transparency mitigates potential biases, promotes innovation and democratizes access, spurring decentralized development. They contend open models foster community-driven improvements and broader application of the technology. Advocates of closed models emphasize the ability to control use cases, implement robust safety and harm reduction measures and provide for sustainable business models. They emphasize closed models enable better regulation of misuses, such

Thomson Reuters is a commercial publisher of content that is general and educational in nature, may not reflect all recent legal developments and may not apply to the specific facts and circumstances of individual transactions and cases. Users should consult with qualified legal counsel before acting on any information published by Thomson Reuters online or in print. Thomson Reuters, its affiliates and their editorial staff are not a law firm, do not represent or advise clients in any matter and are not bound by the professional responsibilities and duties of a legal practitioner. Nothing in this publication should be construed as legal advice or creating an attorney-client relationship. The views expressed in this publication by any contributor are not necessarily those of the publisher.

as the creation of deep fakes and harmful code or applications that perpetuate unlawful or pernicious bias and discrimination.

As the debate between proponents of open and closed models plays out, there are several key legal and operational implications of adopting open genAI to weigh.

Notwithstanding the ideals the OSAID reflects, open models often come with a number of licensing restrictions. These restrictions can limit commercial use, require attribution or impose other conditions that affect the models' use. Legal counsel must review these licensing terms to ensure compliance and to avoid potential legal conflicts, especially when integrating open models into commercial products or services.

Note that inconsistencies in licensing across a model's components and training data can further complicate the use of open models. Distinct licenses may govern different parts of a model — such as the code, pre-trained weights, datasets or a fine-tuned version of the underlying model. This can lead to potential legal conflicts if the licenses are incompatible. For instance, a model might incorporate components that prohibit commercial use alongside others that would allow it.

Users considering whether to rely on open models should consider if their use heightens the risk the user will ultimately bear various unmitigated liabilities, including for copyright infringement.

In addition, while the training data may not materially differ as between open and closed models, the use of open datasets — such as the Books3 dataset, which contains the text of nearly 200,000 books — introduces significant copyright risks.

For example, developers of popular genAI models are facing infringement suits based on allegations they trained models on Books3 without permission from copyright holders. The model developers have argued their use of books or other content to train genAI models constitutes fair use under applicable copyright law. Such ongoing legal battles underscore the inherent risks and complexities of using open datasets.

In response, a number of companies making closed models commercially available have agreed to indemnify users for intellectual property infringement claims based on use of the models or their outputs. Closed models also often include filtering mechanisms to remove content that might infringe copyrights from outputs, and use of such filters is typically a condition of indemnification.

In contrast, open models are generally licensed "as is," without warranties, indemnities or filters, potentially increasing the risk of copyright infringement. Users considering whether to rely on open models should therefore consider if their use heightens the risk the user will ultimately bear various unmitigated liabilities, including for copyright infringement.

That said, open models' licensing terms typically provide (expressly or by default under applicable law) that users own their inputs and outputs, without further obligation to license or otherwise make them available to third parties. While some closed models similarly give users' rights in inputs and outputs, there are often licenses granted back to the vendor that may undermine the rights the user may expect to have in inputs and outputs and create greater risk of intellectual property or confidential information leakage.

Legal counsel need to weigh such trade-offs based on the intended use cases to determine whether it is appropriate to take on incremental copyright risk to retain rights in and confidentiality of inputs and outputs.

In addition to copyright risks, open datasets often contain personal data scraped from the internet, raising significant privacy concerns. Compliance may require the technically challenging tasks of removing personal data or anonymizing or pseudonymizing all data before use.

While both closed and open models may rely on open datasets containing personal data, privacy risks may be lower when using closed models, as the companies offering closed models may be more likely to have proprietary techniques to filter personal data from the training data or the outputs for purposes of their own privacy compliance.

Open models' public access to the source code and model weights may foster collaboration and innovation, but it also introduces security risks. For example, malicious actors can exploit this openness to introduce vulnerabilities or manipulate the model's behavior by altering its weights and publishing the modified versions.

To safeguard the integrity and reliability of these models, organizations should implement robust security measures, ensure models are retrieved from trusted sources and conduct thorough testing and continuous monitoring. While cybersecurity concerns also apply to closed models, the existence of a vendor counterparty to whom risks can be shifted may mitigate the risks to users.

Finally, while open models may allow greater flexibility, including the ability to fine-tune models for specific use cases, they also may introduce additional risks with respect to bias, discrimination and other harms from use of genAI models. Closed models are often deployed only after pre-training, fine-tuning filtering and system prompting aimed to prevent the models from outputting harmful content.

While this is also true of some open models, legal counsel should note that open models are generally static (i.e., the weights are not continually updated and republished), and so harmful behavior of open models or known exploits of open models may not be regularly patched. In addition, open models are often fine-tuned to generate "uncensored" versions that remove many of the guardrails initially built into the open models. Those uncensored models may be able to better fit various use cases, but they also may increase the risk of harmful use.

On the other hand, open models might in theory be more carefully managed and monitored to mitigate harms to an even greater extent than closed models due to the greater transparency of open models.

In summary, use of either closed or open genAI models comes with trade-offs. Legal counsel should therefore carefully track which genAI models are being used and have a thorough understanding of the applicable use cases, the provenance of such models and the applicable licensing terms, establishing clear governance frameworks, conducting continuous education and training and maintaining rigorous monitoring and security measures to help mitigate the risks and legal challenges associated with both open and closed models.

Ken D. Kumayama and Pramode Chiruvolu are regular, joint contributing columnists on legal issues in artificial intelligence for Reuters Legal News and Westlaw Today.

About the authors

Ken D. Kumayama (L) is a partner in the intellectual property and technology group at **Skadden, Arps, Slate, Meagher & Flom LLP**. He concentrates his practice on transactional matters relating to intellectual property, technology, privacy and cybersecurity, as well as artificial intelligence and machine learning. He can be reached at ken.kumayama@skadden.com. **Pramode Chiruvolu** (R) is a counsel in the intellectual property and technology group at the firm. He advises clients on complex transactional matters involving emerging technologies, including artificial intelligence, digital health and biotechnology, the internet of things and 5G networks. He can be reached at pramode.chiruvolu@skadden.com. The authors are based in Palo Alto, California.

This article was first published on Reuters Legal News and Westlaw Today on August 15, 2024.