

Unlocking the Weights: What Enterprises Should Know Before Deploying Open-Weight AI Models

August 12, 2026

Does every task in the enterprise really call for the biggest model on the market? That's the question more companies are now asking. Frontier models set a high standard for pushing the boundaries of what AI can do, but as token fees climb and many everyday tasks turn out not to need frontier-level horsepower, companies are increasingly adding open-weight models to the mix as a complementary option. Deploying them, however, raises legal and governance questions that differ from those raised by hosted AI services or traditional open-source software procurement. This alert is intended to help companies and their general counsel understand and navigate those differences.

Open weight, not open source

"Open weight" is not "open source." Open-source software is generally distributed under standardized, well-understood licenses that grant broad rights to use, modify and redistribute. Open-weight models simply make a model's weights – the numerical settings a model learns during training that shape how it responds – available for download and local deployment. They require integration through a separate software platform to run and manage the model, and often remain subject to bespoke contractual terms on commercial use, intellectual property (IP), redistribution, attribution and downstream deployment.

Key legal risk areas

Organizations considering open-weight models should evaluate the legal framework governing deployment, not just technical performance. Licensing, IP, privacy, security and an evolving regulatory landscape all affect how these models can be used and what safeguards should accompany their deployment.

A. Licensing is the threshold issue

For many organizations, the first legal issue in evaluating an open-weight model is not copyright or regulation, but the license itself. Unlike traditional open-source licenses, open-weight licenses are increasingly customized. They typically permit downloading, deploying and modifying weights, but may impose obligations well beyond preserving copyright notices or attribution.

Depending on the model, organizations may encounter provisions addressing commercial use limitations, acceptable use restrictions, volume restrictions, and attribution or branding requirements. Some licenses also restrict using the model, its outputs or derivative works to train, develop or improve a competing AI model. Like traditional open-source licenses, organizations may also occasionally encounter "copyleft" provisions requiring that derivatives thereof be made available on the same open-source terms, which may conflict with commercial objectives of keeping the organization's software source code a trade secret. The same issue can arise with training datasets, such as those licensed under the Creative Commons ShareAlike license, which requires that new works built from the same data be shared under the same license. That said, this concept has traditionally applied to creative works that are direct derivatives of other creative works and how it applies to models trained on those works remains a fact-specific analysis.

Intended deployment matters too. Internal productivity use may carry different obligations than incorporating the model into a customer-facing product or platform. Organizations planning to fine-tune, redistribute derivative weights or build downstream products should confirm that the license permits this, and check for added obligations, taking into account how their product's use may evolve.

B. IP questions continue to evolve

Open-weight models also raise IP questions that courts and regulators are still working through as generative AI develops. Much of the current litigation asks whether using copyrighted material for AI training infringes copyright or falls within doctrines such as fair use. These disputes generally involve developers rather than downstream deployers, but organizations should recognize that training data provenance is not always transparent, and the governing standards remain unsettled. Questions also arise over AI-generated outputs: ownership of generated content, resemblance to protected third-party works, risk of incorporation of open-source code and security vulnerabilities into AI-generated software, and contractual allocation of IP risk.

Unlike many hosted AI services, self-hosted open-weight deployments may lack provider indemnification or related contractual and technical protections – not only for IP-infringing outputs, but also more broadly for harmful, inaccurate or discriminatory outputs. Organizations should therefore build governance around AI-generated outputs, including human review, documentation, and technical validation and safeguards, and consider how liability for AI-related risks is addressed when a model provider offers no contractual indemnities or other protections.

C. Privacy, security and deployment architecture

A principal advantage of open-weight models is deployment flexibility. Unlike provider-hosted models, which require transmitting prompts and data to a third party, open-weight models can run entirely within enterprise-controlled environments – a meaningful benefit for organizations handling sensitive commercial information, proprietary IP, or regulated data subject to sector-specific privacy and data-handling rules. That flexibility comes with a trade-off: It shifts responsibility for securing and operating the AI environment to the deploying organization. That includes integration, access controls, infrastructure maintenance, usage monitoring, safeguards against inappropriate content, vulnerability management, output optimization, and compliance with privacy, breach-notification and cybersecurity requirements that a hosted provider's data processing terms might otherwise help address.

Conversely, managed inference providers – companies that host and run a model on their own infrastructure so customers can access it without operating it themselves – may offer contractual protections, support and established security controls, but they introduce their own vendor-management and data privacy-governance considerations. The right approach depends on the use case, legal obligations, risk tolerance and governance capability, not just technical requirements.

D. Regulatory landscape continues to develop

The legal framework for advanced AI continues to evolve, as lawmakers weigh AI governance, export controls, national security, computing restrictions, consumer protection and cross-border deployment. Many regulatory frameworks do not yet distinguish open-weight from proprietary models. One notable exception is the European Union AI Act, which exempts open-source general-purpose AI model providers from certain technical documentation and downstream information obligations. It does not exempt them from all requirements; they must still implement a policy to respect EU copyright law and rightsholders' text and data mining opt-outs and publish a sufficiently detailed public summary of the content used for training. Recent reporting likewise suggests that the White House's voluntary, nonpublic frontier model review guidelines do not apply to open-weight models – further evidence that this remains an evolving area of AI governance and national security policy.

Adding to this complexity, a substantial share of today's top open-weight models are developed outside the United States, including by developers based in China. That reality has drawn its own share of policy attention. The US and other governments have begun considering supply-chain provenance, data-handling practices and security review as part of a broader conversation about foreign-developed AI models generally, separate and apart from the quality or utility of any particular model. Organizations evaluating an open-weight model of foreign origin should treat these considerations as part of standard diligence – alongside licensing and IP review – rather than as a bar to adoption, while staying alert to guidance that may apply specifically to models associated with certain jurisdictions or certain entities.

Open-weight model deployments may therefore raise a broader, and in some respects more nuanced, range of legal considerations than traditional software procurement, and even hosted frontier models.

Lawmakers have also begun examining the cross-border movement of advanced AI models, weights and related technology. Organizations should monitor developments affecting cross-border deployment and jurisdiction- or provider-specific restrictions, which remain highly dynamic and may affect deployers and developers.

Pre-deployment checklist

Before deploying an open-weight model, organizations should, at a minimum:

- Carefully review the applicable license to confirm it permits the intended deployment, and identify any restrictions on commercial use, redistribution, fine-tuning, downstream deployment, attribution or other contractual obligations that may apply.
- Assess the jurisdiction, ownership and supply chain associated with the model's developer and any upstream contributors, including where the model was trained and hosted; confirm whether the developer or model is subject to export control classification, entity-list or other trade restrictions; and where the deployment involves sensitive, regulated or government-related data, consider whether additional national security or cross-border review is warranted before proceeding.
- Evaluate whether a self-hosted or managed inference architecture is more appropriate given the organization's data sensitivity, operational needs, risk tolerance, customer commitments and available contractual protections.
- To the extent possible, diligence the provenance and licensing history of any third-party model weights before deployment, particularly where the model has been modified, fine-tuned or obtained through an intermediary, to understand what rights and obligations accompany the model.
- Implement governance around AI-generated outputs, including human review, technical safeguards and code-scanning to help catch and prevent potential IP issues, open-source software, security vulnerabilities and other material errors before outputs are deployed or relied upon.
- Build a process to keep policies current as fast-moving export controls, AI regulation and other legal rules continue to evolve.
- Develop contingency plans that account for changes to licensing terms, model availability or regulatory requirements that could affect continued deployment or commercial use.

As enterprise adoption of open-weight models grows, the legal questions surrounding deployment – from bespoke licenses to IP risk, governance and emerging regulation –will keep evolving. Organizations that evaluate these issues early in procurement and deployment will be better positioned to fold open-weight models into their AI strategies while managing legal and operational risk.

Cooley combines experience in AI, technology transactions, IP, privacy, cybersecurity, national security and global AI regulation to help clients evaluate, deploy and govern AI systems across the company.

This content is provided for general informational purposes only, and your access or use of the content does not create an attorney-client relationship between you or your organization and Cooley LLP, Cooley (UK) LLP, or any other affiliated practice or entity (collectively referred to as "Cooley"). By accessing this content, you agree that the information provided does not constitute legal or other professional advice. This content is not a substitute for obtaining legal advice from a qualified attorney licensed in your jurisdiction, and you should not act or refrain from acting based on this content. This content may be changed without notice. It is not guaranteed to be complete, correct or up to date, and it may not reflect the most current legal developments. Prior results do not guarantee a similar outcome. Do not send any confidential information to Cooley, as we do not have any duty to keep any information you provide to us confidential. When advising companies, our attorney-client relationship is with the company, not with any individual. This content may have been generated with the assistance of artificial intelligence (AI) in accordance with our AI Principles, may be considered Attorney Advertising and is subject to our [legal notices](#).

Key Contacts

Sean Quinn New York	squinn@cooley.com +1 202 728 7075
------------------------	--------------------------------------

William K. Pao Los Angeles	wpao@cooley.com
Janet H. Kim Washington, DC	janetkim@cooley.com +1 202 728 7060
Tracy Rubin Palo Alto	trubin@cooley.com +1 650 843 5568
Michael Egan Washington, DC	megan@cooley.com +1 202 776 2249
Brian E. Nelson Washington, DC	brian.nelson@cooley.com +1 202 728 7152
Lorenzo Sordi New York	lsordi@cooley.com +1 212 479 6601
Adam Silow New York	asilow@cooley.com +1 212 479 6163
Ellie Dupler Los Angeles	edupler@cooley.com +1 213 698 6042
Isamar Vaquero Washington, DC	ivaquero@cooley.com +1 202 776 2217
David O'Connor London	doconnor@cooley.com +44 20 7556 4137

This information is a general description of the law; it is not intended to provide specific legal advice nor is it intended to create an attorney-client relationship with Cooley LLP. Before taking any action on this information you should seek professional counsel.

Copyright © 2023 Cooley LLP, 3175 Hanover Street, Palo Alto, CA 94304; Cooley (UK) LLP, 22 Bishopsgate, London, UK EC2N 4BQ. Permission is granted to make and redistribute, without charge, copies of this entire document provided that such copies are complete and unaltered and identify Cooley LLP as the author. All other rights reserved.

