AI PROGRESS

June 9, 2026

Dear Members of the Judiciary Committee,

AI Progress is a 501(c)(4) organization dedicated to the responsible development of artificial intelligence (AI) tools. The coalition emphasizes existing U.S. copyright law's essential role in fostering innovation, protecting intellectual property rights, and securing U.S. competitiveness in AI innovation. Members include Amazon, Anthropic, Cohere, Google, Meta, Midjourney, Microsoft, and OpenAI.

We write to express our profound concerns regarding Assembly Bill 412 (AB 412), the AI Copyright Transparency Act. This legislation would impose impractical and unworkable requirements on AI developers, ultimately stifling technological innovation and compromising both California's and the United States' leadership in artificial intelligence (AI).

Fundamental Compliance Challenges

The bill's core requirement — forcing AI developers to document and identify registered copyrights and their owners for all training materials — is fundamentally not feasible.

Hundreds of thousands of works are registered with the U.S. Copyright Office (USCO) every year. This means AI developers would need to examine billions of individual works,[1] determine whether they are registered with USCO, and if so, identify the current copyright owners. This is not feasible given there is no comprehensive, reliable and up-to-date machine-readable database of copyrighted works to facilitate such assessments.

Compounding this lack of any reliable copyright metadata are the additional complications that copyright ownership may change (and not be updated with USCO), copyrighted works may be jointly created and owned, and different types of works may form a single "work." A digital file of a song, for example, may include multiple copyrighted works (and owners), including the underlying musical composition and lyrics, which are protected as a literary work, versus the sound recording of the composition's performance. Moreover, just the musical composition alone can be owned by multiple individuals and/or entities.

Additionally, large language models are trained on vast amounts of web-crawled data, making comprehensive copyright tracking virtually impossible. The sheer volume and complexity of these datasets present insurmountable obstacles that render compliance not just difficult but practically unachievable in the current legal and technological landscape.

While we acknowledge recent amendments have narrowed the bill's scope, significant challenges remain. Copyright ownership is frequently contested; many works contain multiple copyrightable

[1] One example: Common Crawl, a regularly updated open-source collection of scraped web data, consists of more than 250 billion web pages (https://commoncrawl.org/). This dataset is frequently used by AI developers for training models.

elements (text, images, audio, video), and, in some cases, owners make their works available for a variety of purposes (see Creative Commons licenses). Works created before 1978 and anonymous works further complicate attribution. AB 412 offers no clear mechanism for dispute resolution or for handling ownership uncertainties, exposing companies to litigation risk.

Given that no definitive source of copyright ownership information exists, meaningful compliance is unattainable. The bill's compliance burden is staggering — tracking and verifying registered copyrighted materials at the scale required would impose prohibitive costs. Moreover, the bill's seven-day compliance window for responding to data requests is wholly unrealistic, especially for small businesses and startups.

AB 412 would also introduce new, unnecessary statutory damages for noncompliance. Given the impossibility of meaningful compliance discussed above, the bill's statutory damages structure is especially problematic. It would penalize AI developers for inadvertent errors in matching every piece of content to its rightful copyright owner, but it is ambiguous on what constitutes a "violation" and whether inadvertent errors or omissions would be treated as separate infractions. If every mistake is counted as a distinct violation, compliance would become even more costly and burdensome, deterring AI research and innovation in California.

Unreliability of Fingerprinting

The bill suggests AI developers implement reliable fingerprinting techniques to track copyrighted materials. However, fingerprinting methods are inherently statistical and prone to false positives, introducing further legal and operational uncertainty.

Additionally, training data undergoes significant transformation. Training data is transformed into numerical representations embedded within complex parameter networks, making effective fingerprinting not feasible.

The final datasets used are typically reformatted and significantly modified to optimize them for machine learning. This data preparation process changes the fundamental structure that an "approximate fingerprinting" process would be looking for. Attempting to fingerprint data before this preparation stage risks flagging vast amounts of data that may not even be used in the final training process.

A key challenge is that the more often similar data appears within a training dataset (which is common in large-scale datasets), the higher the likelihood of false positives being identified.

Impact on AI Training

Existing U.S. copyright law, including the well-established fair use doctrine, provides a framework that facilitates AI training while protecting intellectual property rights. AB 412 threatens to disrupt this framework, creating unnecessary legal and operational barriers that would harm innovation and competitiveness.

Innovation in AI fundamentally depends on the ability of AI models to be trained on large quantities of data. The technical process of "learning" for AI models means being able to derive patterns, structures,

and relationships from across a broad body of content so that they can operate on probabilistic modeling. The amount of data used in training today's AI models improves accuracy and robustness.

Access to training material is crucial for creating effective and accurate AI models. But the burdens imposed by AB 412 would severely impact the ability of U.S.-based AI developers — many of whom are headquartered or operate in California — to train advanced models. This would place the competitiveness of American-made AI systems at a disadvantage in the global AI landscape.

If enacted, AB 412 will impede the use of data to train AI models, ultimately undermining the quality of American-made AI models. In this way, AB 412, would directly undermine the development of AI-driven technologies that can solve real-world problems, solve societal challenges, drive scientific discovery, and create new opportunities for growth and innovation across industries.

Risks to Privacy and Trade Secret Information

Beyond its compliance challenges, AB 412 introduces significant risks regarding privacy and proprietary information. Section 3117 of the bill creates a mechanism that could be exploited by bad actors to access proprietary data or sensitive personal information. The bill lacks adequate safeguards to prevent the misuse of information requests, potentially exposing California-based AI companies to unfair competition from international rivals.

By forcing AI developers to disclose extensive details about their training datasets, AB 412 increases the likelihood of exposing trade secrets without a clear verification process for copyright ownership claims. The absence of protections against frivolous or bad-faith information requests further compounds these risks, undermining AI companies operating in California and exposing companies to increased litigation risks.

Federal Preemption and Existing State Legislation

AB 412 conflicts with the Constitution's Supremacy Clause, raising concerns regarding federal preemption. Copyright law falls under federal jurisdiction, and AB 412's additional enforcement obligations would create additional legal uncertainty at a time when AI-related copyright issues are already under judicial review. Ultimately, this will require litigation to resolve, which will be costly and burdensome for both AI developers and the State.

Conclusion

Given AB 412's practical challenges, conflicts with federal law, and detrimental impact on AI development, we urge the Committee to oppose AB 412. The bill's significant compliance burdens and legal uncertainties would have a negative impact on California businesses and, as a consequence, America's global AI competitiveness.

Sincerely,

AI Progress