

BSA Comments on the Guidelines for Ensuring the Appropriateness of Research, Development, and Utilization of AI-Related Technologies (Draft)

December 11, 2025

relevant section	Page #	Line #	comments
Overall			<p>The Business Software Alliance (BSA) ⁽¹⁾ appreciates the opportunity to comment on the draft Guidelines for Ensuring the Appropriateness of Research, Development, and Utilization of AI-Related Technologies (“Guidelines”). BSA is the leading advocate for the global software industry, and our members are enterprise software companies that lead in AI, cloud storage services, cybersecurity solutions, quantum computing, and other breakthrough technologies. These companies support businesses in every sector of the economy, providing trusted tools that help other businesses innovate and grow. We hope the following recommendations will support policies to realize “trustworthy AI” in Japan.</p> <p>To enable meaningful, substantial input from relevant stakeholders, we strongly recommend the Government to allow sufficient time to comment for at least three to four weeks. Having English translation of the draft documents prepared by the</p>

			<p>Government is particularly helpful when words are used that leaves room for several interpretations.</p> <p>(1) https://www.bsa.org/ (global site) / https://bsa.or.jp/ (Japanese site)</p>
<p>2 Key Matters Research & Development Institutions and Utilizing Business Operators Should Address (2) Ensuring Transparency to Build Trust with Stakeholders</p>	<p>Page 7 of English translation/ Page 6 Page 7 of original Japanese draft</p>	<p>20-17 of English translation/ 16-18, 1-4 of original Japanese draft</p>	<p>The Guidelines describes AI business operators as “utilizing business operators developing and providing products and services using AI.” While we understand this is in line with the definition of Article 7 of the AI Act (2), it is important to recognize that responsibilities should be feasible to implement and fit the roles of different actors in the AI ecosystem, including developers, integrators, deployers, as different organizations will have varying access to relevant information and different abilities to address issues that may arise.</p> <p>Also, the Guidelines encourage AI utilizing business operators to ensure transparency to stakeholders and provide information within a reasonable scope on the source of training data and the generated outputs. While we understand that this is presented as voluntary measures, as recognized in footnote 17 of the Guidelines, when companies are required to disclose detailed information about how they collect and curate the datasets used to train AI models, it can threaten their ability to protect important trade secrets. For example, disclosing specific datasets an AI model is trained on and how that data is curated can require a company to disclose how it trains its models. The process of training and fine tuning models involves significant research and development. Not only would this disclose trade secret information, it may also give bad actors a potential roadmap for future cyberattacks.</p> <p>Also, certain disclosure proposals could be impossible to implement, such as those that would require AI developers to disclose all copyrights, and all copyright holders indirectly associated with an AI training dataset. Given that copyright arises upon creation, everything from online restaurant reviews to blogs to digital recordings may be potentially subject to copyright. However, the provenance, duration, ownership, and/ or any licensing details are neither self-evident nor reasonably knowable. For example, a single work may involve</p>

			<p>multiple overlapping rights holders, or “orphan” works with no clear claims. When developing or using AI at scale, it is unworkable to make granular determinations whether any part of a tokenized dataset used in AI training comes from part of a copyrighted work. Moreover, these sorts of copyright-based disclosures are in tension with exceptions to copyright protections, that generally cover AI training. Requiring AI operators to publicly disclose their training datasets lacks a sound legal or practical foundation. Such data often comprises proprietary collections—curated, licensed, or generated through substantial commercial investment—and as stated earlier, may contain trade secrets essential to competitive differentiation.</p> <p>We therefore highlight the importance of distinguishing between disclosures that are appropriately made to regulators, to users, or to the public. Focus on disclosures of high-risk uses of AI tools, rather than low risk uses. Disclosure requirements should focus on high-level information that helps consumers understand the capabilities of AI tools, without requiring granular detail about datasets used for training.</p>
<p>2 Key Matters Research & Development Institutions and Utilizing Business Operators Should Address (3) Ensuring Sufficient Safety</p>	<p>Page 8 of English translation/ Page 7 of original Japanese draft</p>	<p>1-12 of English translation/ 6-14 of original Japanese draft</p>	<p>To reduce the risks of the spread of misinformation/disinformation, the Guidelines could further be improved by referencing existing standards such as the Coalition for Content Provenance and Authenticity (C2PA) for content authenticity and provenance. This standard will help consumers decide what content is trustworthy and promote transparency around the use of AI.</p>



<p>2 Key Matters Research & Development Institutions and Utilizing Business Operators Should Address (5) Consideration for Stakeholders, Recognizing the Importance of Data as the Foundation for AI Innovation</p>	<p>Page 8 of English translation/ Page 7 Page 8 of original Japanese draft</p>	<p>21-25 of English Translation/1-5 of original Japanese draft</p>	<p>The Guidelines encourage businesses developing and providing AI with significant social impact to strive to consider and implement measures for building an ecosystem that returns profits to data holders, such as intellectual property owners. While we recognize the importance of ensuring that all stakeholders benefit from AI, it would not be appropriate to require licensing for training on publicly available or otherwise legally accessed content, since this is not a copyright infringement. The current copyright system in Japan enables innovation, growth and development of new AI use cases that can support content creation, distribution and underpin new business models for content creators. The current copyright system in Japan can therefore provide profits to data holders without introducing a licensing regime which would stifle innovation.</p> <p>Further, expecting direct licensing to AI operators, even for publicly available content, presupposes that an AI developer can identify and negotiate with every relevant rightsholder. In many real-world contexts, however, it is impossible to determine who owns the rights to the materials that appear online. For example, an entity in Japan might wish to train an AI model to detect misinformation patterns or abusive language in online media. The data needed for such a project would include comments on social media and forum discussions — all of which may contain copyrighted material. Determining the ownership in such scenarios would be potentially cost prohibitive: some posts may belong to individual users, others to the hosting platform under its terms of service, and others still may have been copied or quoted from yet another source. Furthermore, the total contribution of any particular piece of content, among the potentially billions of sources of data used to train larger models, would be miniscule.</p> <p>To any government that may consider mandating licensing, BSA cautions the substantial chilling effect this will have on innovation. Requiring new market</p>
---	--	--	---

			<p>entrants or less resourced entities to secure permission for each item of content used in model training would make it virtually impossible to undertake socially beneficial machine learning projects. The transaction costs and administrative burden of identifying and negotiating with thousands of disparate rightsholders would far outweigh the value of the research itself. It would also inevitably concentrate AI capability in the hands of existing market participants and entities with the financial capabilities to absorb the licensing costs, creating an uneven playing field in AI development. This would marginalize new entrants and stifle open-source AI development.</p>
--	--	--	---

