January 26, 2026

BSA COMMENTS ON THE DRAFT PRINCIPLE-CODE FOR THE PROTECTION OF INTELLECTUAL PROPERTY AND TRANSPARENCY FOR THE APPROPRIATE USE OF GENERATIVE AI

General Comment

[Category of Comment (Relevant Section): (3) Others]

The Business Software Alliance (**BSA**)[1] appreciates the opportunity to comment on the Draft Principle-Code for Protection of Intellectual Property and Transparency for the Appropriate Use of Generative AI **(draft Code**).[2] BSA is the leading advocate for the global enterprise software industry, and our members create technology solutions that power other businesses, including artificial Intelligence (**AI**), cloud storage services, cybersecurity solutions, quantum computing, and other breakthrough technologies. Our members are global leaders in developing, tailoring, integrating, and deploying AI systems and services, and the tools used by others in the development of AI systems and applications. As a result, they have unique insights into technology's tremendous potential to spur digital transformation and the policies that can best support the responsible use of AI.

BSA recognizes that the intersection of AI and intellectual property rights is a sensitive and evolving issue, both in Japan and around the world. The creative industries are an important component of Japan's economy and contribute to its cultural identity. It is important to collaboratively examine policies around intellectual property that fosters the responsible use of AI in ways that stimulate the revitalization of rightsholders, artists, and content creators, including software developers, while continuing to protect rights. It is also equally important to ensure these policies do not inadvertently inhibit Japan's ability to harness AI to enhance productivity, drive innovation across sectors, and strengthen its global competitiveness as Japan moves toward becoming the "most AI-friendly country in the world (i.e. the country that is the easiest to develop and utilize AI)."

[1] The Business Software Alliance (www.bsa.org) works in over 20 markets in the US, Europe, and Asia, advocating for policies that build trust in technology so that every industry sector and the public can benefit from innovation.

BSA's members include: Adobe, Alteryx, Amadeus, Amazon Web Services, Asana, Atlassian, Autodesk, Avalara, Bentley Systems, Box, Cisco, Cloudflare, Cohere, Cohesity, Dassault Systemes, Databricks, Docusign, Dropbox, Elastic, EY, Graphisoft, HubSpot, IBM, Informatica, Kyndryl, MathWorks, Microsoft, Notion, Okta, OpenAI, Oracle, PagerDuty, Palo Alto Networks, PTC, Rubrik, Salesforce, SAP, ServiceNow, Shopify Inc., Siemens Industry Software Inc., Trend Micro, TriNet, Veeam, Workday, Zendesk, and Zoom Communications Inc.

[2] https://www.kantei.go.jp/jp/singi/titeki2/pdf/shiryo3.pdf

Recommendations

[Category of Comment (Relevant Section): (2) Principle and Exception Presented in the Document]

Ensuring Legal Certainty and Innovation-Friendly Implementation of AI Transparency Measures

BSA member companies are strongly committed to advancing innovation and protecting intellectual property, and they recognize the importance of promoting transparency. The purpose of transparency should be about ensuring that downstream users of AI systems can understand the capabilities and limitations of a model and are able to make informed choices when using the model. However, the draft Code's broad scope of application and its stringent approach raise several concerns.

Although the draft Code characterizes compliance by AI businesses as voluntary, it also provides that compliant companies will be publicly listed, together with links to their corporate websites displaying disclosed information, and that companies will be expected to explain their reasons for non-compliance. It also indicates that the Government intends to evaluate the compliance status of AI businesses, providing certain incentives in public sector initiatives. This structure risks creating de facto obligations despite the draft Code's voluntary framing.

The draft Code further states that generative AI businesses are not required to forcibly disclose information belonging to them, including trade secrets. At the same time, however, it expresses the Government's expectation that the cumulative efforts of individual AI businesses will result in the standardization of disclosed information. In practice, this approach could impose significant constraints on corporate activities and may undermine the effectiveness of the Japanese Copyright Act in promoting AI development.

It is important to clarify that there are no legal requirements under the Copyright Act or any other regulations in Japan that require AI model developers or AI service providers to disclose information relating to AI development or training. In particular, Article 30-4 of the Japanese Copyright Act explicitly permits the reproduction and other uses of works to the extent necessary for information analysis, provided the purpose is not for enjoying the expressive content, which include the extraction, comparison, classification, and statistical processing of large volumes of data. This exception reflects a deliberate policy choice by the Japanese legislature to ensure that innovation in fields such as AI, machine learning, and data science is not impeded by legal uncertainty.

AI businesses reasonably expect that, once they have ensured compliance with applicable laws, they will not be subject to additional requirements or additional voluntary commitments. The "comply or explain" approach adopted in the draft Code risks creating legal uncertainty for AI businesses and runs counter to Japan's policy of supporting AI innovation.

Recommendations: The Intellectual Property (**IP**) Strategy Headquarters (**HQ**) should amend Principle 1 to delete language, such as "use of data to develop and train generative AI models should not infringe on intellectual property rights" and provisions that require AI businesses to "provide contact points for

relief for rightsholders," that imply training on copyrighted material may itself give rise to copyright infringement enforcement actions or amend these provisions to clearly state that they are not intended to apply to lawful training activities.

The IP Strategy HQ should explicitly state that the draft Codes are positioned very clearly as legally non-binding reference Guidelines.

Concerns Regarding Overbroad Scope and Public Disclosure Commitments

Although the draft Code explains that it draws on measures adopted under the EU AI Act and the Stewardship Code in the area of corporate governance, the draft Code's scope is considerably broader. The draft Code applies to all generative AI developers and providers offering AI products and services to the public, rather than being limited to foundation model developers or entities presenting systemic risk, as was the case for the July 2025 General-Purpose AI Code-of Practice in EU. In addition, the draft Code emphasizes public-facing transparency, rather than transparency toward regulators or downstream AI providers, which depart from the EU's risk-based and tiered approach.

AI businesses invest substantial resources in developing their training methodologies and AI development processes which they regard as confidential and commercially sensitive trade secrets. Information relating to AI model development is highly sensitive and public disclosure of such information would materially harm the legitimate interests of AI businesses. Moreover, given the rapid pace of technological change and the evolving state of AI development, ongoing or frequent disclosure requirements risk unintentionally resulting in the dissemination of inaccurate, incomplete, or outdated information.

Asking AI businesses to disclose specific information relating to AI training, such as detailed training methodologies and granular information about training data, for example, listing URLs, goes beyond what is necessary to provide transparency and would not serve the legitimate interests of rightsholders or other stakeholders.

An AI model does not store or retrieve copies of the training data; instead the model learns patterns and generalizes features across the corpus of training data. It is not technically possible to trace a specific output back to any single source or URL. Since an AI model only analyzes statistical patterns, AI training (input-side analysis) is neither technically nor legally relevant to determining copyright infringement. Any similarity in outputs is likely to result from commonly occurring features, such as well-known characters that recur in a variety of training data rather than from any specific work used in training. Input-side analysis is legally irrelevant because AI training — i.e., computational data analysis — does not involve the consumption of any copyrighted works for their expressive content, as explained earlier.

Since disclosures are neither technically nor legally relevant to determining copyright infringement, they risk exposing commercially sensitive information, undermining investment in research and development, and discouraging AI model developers from making their models available in Japan. Also, while we support efforts to avoid using content from pirated sites, any obligation should be feasible and avoid censoring

information online. It is important to recognize that AI developers already take steps to avoid using content from pirated sites to train their models. Promoting voluntary best practices to avoid using content from pirate websites is a positive suggestion, especially with respect to websites that have been found to make available to the public – persistently, repeatedly, and on a commercial scale – unauthorized copies of copyright-protected works in contravention of Japan's copyright law.

In addition, copyright is a private right held by authors and the interests protected under the Unfair Competition Prevention Act similarly constitute private rights belonging to specific individuals or corporations. Disputes concerning these rights should be resolved between the relevant parties.

Government authorities therefore should refrain from intervening in a manner that effectively advantages one party in a potential dispute by compelling AI developers to disclose proprietary or confidential information. AI developers and rights holders already have established mechanisms for direct engagement, including commercial licensing negotiations.

Recommendations: The IP Strategy HQ should amend the draft Code to explicitly limit transparency commitments to high level, non-confidential information, and to further clarify that the disclosure of trade secrets, confidential contractual information, or confidential business and technical information is not expected.

The IP Strategy HQ should eliminate the creation of additional disclosure powers or procedures in the draft Code that may create advantages or disadvantages to the relevant parties.

Safeguarding Cybersecurity and National Security within Transparency Obligations

Furthermore, we urge the Government to consider the serious potential national security impacts when formulating the details of disclosure requirements. Public disclosures that require information detailing how network defenders use and train AI systems to secure networks could unintentionally create a roadmap for cyber adversaries to circumvent those defenses, in turn jeopardizing the underlying security of network and information systems.

The requirement of technical disclosures in Principle 1, such as "details of the model training process" and "parameter settings" presents a significant risk when applied to AI used for cybersecurity reinforcement. This level of detail could provide attackers with a blueprint of the defense architecture. For instance, if a threat actor understands how a cybersecurity AI model is tuned to detect malware, they can modify their attack vectors to bypass detection. Therefore, it is especially important to ensure that cybersecurity enabling AI systems are exempt from granular architectural disclosures.

Further, with regards to Principle 2, which urges AI businesses to provide data access for legal action, it is important to recognize that training data used for cybersecurity AI often includes sensitive threat intelligence and proprietary samples. Allowing third parties to access this data merely because they are "preparing to take legal action" risks inviting abusive fishing expeditions by malicious actors seeking access to sensitive security intelligence. Access to such sensitive data should be subject to a much higher legal threshold.

Similarly, Principle 3, which requests AI businesses to provide URLs of training data, raises significant security concerns. Cybersecurity companies utilize massive amounts of URL data to train web filters and firewalls. Disclosing whether a specific URL was used in training could inadvertently signal to malicious actors that their infrastructure has been flagged by security systems, allowing them to adapt their infrastructure to evade defenses.

Recommendation: Consistent with our recommendations above, the IP Strategy HQ should revise Principles 1-3 to ensure that cybersecurity enabling AI systems are not expected to disclose internal parameter settings and training processes that could reveal defense logic, limiting disclosure to high-level capabilities, establish a higher threshold, such as a court order, for accessing training data related to security systems to prevent abuse, and ensure such systems are not expected to disclose specific URLs used in training cybersecurity models to protect threat intelligence sources and methods.

Avoid Introducing Disproportionate and Unworkable Commitment

We also want to highlight that Principles 2 and 3 of the draft Code introduce additional compliance expectations that are not contemplated under the EU AI Act. These principles seek commitments from AI businesses to respond to disclosure requests from parties who are taking, or preparing to take, legal action, as well as from generative AI users. Both Principles require AI businesses to provide information regarding data used for training and validation, including non-public datasets obtained through web crawling or third parties, public datasets, and synthetic data. Principle 2 applies in the context of actual or anticipated legal proceedings, while Principle 3 applies to users seeking information about whether training or validation data contains content identical or similar to their generated outputs. Taken together, these Principles represent a significant expansion of commitments that is neither proportionate nor feasible to implement.

The problems with the mandates in both Principle 2 and Principle 3 stem from an inappropriate linkage of training data (e.g., input) with the generated output from generative AI models and a consequent focus on "input-side" analysis rather than the appropriate focus on "output-side" analysis for claims of copyright infringement. The draft Code conflates output issues as being solved through input solutions when, in reality, as stated earlier, the inputs are not relevant, and if there is a concern about an infringing output, they would already have recourse under current copyright law.

A large-scale AI training corpus may contain billions of individual data elements drawn from a vast array of digital sources. As explained earlier, the computational data analysis involved in tokenizing and analyzing such data for training does not involve the consumption of copyrighted works for their expressive content. Rather, such analysis involves mathematical calculations of probabilities, correlations, trends, and other patterns across the entire tokenized data set. Such analysis seeks to understand only the statistical patterns (e.g., the relationships of specific tokens in relation to other tokens) distributed across the entire data set. These statistical patterns are themselves not expressive content protected by copyright law. Nor

does copyright protection extend to facts, ideas, or mathematical concepts, which are the subject of these computational and statistical analytical processes.

Given these facts and the additional important factor of Japan's existing and farsighted computational data analysis exception, requiring or expecting the detailed disclosure of underlying training data to assess concerns about copyright infringement is both impractical and legally irrelevant.

In the case of Principle 2, requiring AI businesses to disclose information to parties merely "contemplating" legal action would set a troubling precedent by effectively bypassing established judicial processes to assess complex legal claims that properly falls within the jurisdiction of the courts. The broad definition of eligible requesters and low threshold for making a request exposes AI developers to a high risk of speculative and abusive requests for disclosure. This would allow claimants to demand access to confidential information, risking disclosure of trade secrets and diverting resources away from research and development. Japan's existing Code of Civil Procedure already provides mechanisms for obtaining relevant information in the context of litigation. Before contemplating any additional measures, it is essential to assess whether existing judicial processes are already sufficient to address legitimate disclosure needs.

Including the mandates under Principle 3, where an AI user seeks specific information to assess copyright infringement in AI generated output may demand disclosure of specific training data, exacerbating the technical and legal infeasibility of these sections of the draft Code.

As mentioned before, AI training input is neither technically nor legally relevant to determining copyright infringement, and therefore, these technically infeasible requirements are simply not necessary to assess copyright law matters. Where AI output bears some similarity to a work protected by copyright, questions of liability for copyright infringement can be resolved on the basis of an output-side analysis alone.

For these reasons and others, it is vastly preferable to focus on AI outputs. As a legal matter, BSA strongly agrees that copyright holders should have full and effective remedies when their rights are infringed by particular outputs. This principle applies equally to outputs generated using AI systems and outputs generated in other ways. As such, existing copyright law should prove adequate to address questions of infringement.

In fact, in cases in which a user discovers that an AI generated output appears similar to content available online, that user already has access to the relevant information necessary to assess similarity and the circumstances under which the output was generated, including their own prompts and iterative creative process.

The legal question of whether content that is similar constitutes copyright infringement is complex and fact dependent. As reflected in existing copyright law doctrines, similarity alone does not establish intentional copying. For example, in many cases a work, such as an artistic image, may be iteratively created using multiple prompts and other methods to build and create the image using the AI tool. In any case, such AI output-side analysis (rather than input-side analysis) comprises the proper focal point for AI and copyright discussions. Relevant output-side questions include: (1) what is the degree of similarity between a

copyright and a generative AI output, (2) is there evidence of independent creation, (3) who generated the output, (4) under what circumstances, and (5) with what intent? All of these questions can be adequately addressed by examining AI outputs in light of existing copyright legal norms.

Finally, focusing on AI output-based solutions offers a pathway to address instances of copyright infringing AI generated output. Many AI developers already implement technical and contractual measures, such as output filtering, meta prompts, and contractual terms, to mitigate the small risk of AI tools generating substantially similar copies of protected works. These measures are appropriately aligned with preventing infringing outputs.

We are concerned that Principles 2 and 3, with their focus on input-side disclosure requirements, will encourage damaging and costly frivolous litigation that will unnecessarily undermine Japan's AI development goals with little actual benefit to rightsholders and their interests.

For all these reasons, there is no reasonable or practical basis for imposing commitments on AI businesses to respond to URL or source-specific inquiries under Principles 2 and 3. These Principles fail to reflect the technical realities of model training, will promote frivolous litigation, and focus on the wrong side of the AI input / AI output divide. As a legal, factual, and technical matter, focusing on AI inputs — rather than AI outputs — will simply not provide a workable or effective solution to copyright-related concerns.

Recommendation: The IP Strategy HQ should reconsider the overall structure and framework of the draft Code and delete Principles 2 and 3 from the draft Code entirely.

Reaffirming Risk-Based, Internationally Aligned AI Policymaking

In developing AI policies, including the draft Code, we strongly encourage the Government to engage in thorough and meaningful consultations with affected stakeholders, including AI businesses and rights holders, consistent with the internationally recognized approach to AI governance that Japan has led through the G7 Hiroshima AI Process (HAIP). Unfortunately, the draft Code adopts an approach that departs from the internationally aligned, risk-based principles underpinning HAIP.

We are concerned that the draft Code would undermine or conflict with ongoing international standardization efforts by copyright holders and AI developers to develop new technological solutions for copyright holders to express their preferences regarding AI training that may involve copyrighted works. At the present time, there are active cross-industry consultations relating to automated tools to indicate a rights-holder does not want a website used for training purposes, similar to the current "do not crawl" tools that apply to search engines. These discussions are occurring in organizations like the Internet Engineering Task Force (IETF), World Wide Web Consortium (W3C), and Coalition for Content Provenance and Authenticity (C2PA). Participants in these and similar discussions include representatives of the creative and technology industries, as well as civil society and academia.

Recommendation: The IP Strategy HQ should refrain from imposing commitments or expectations on AI businesses that could inadvertently undermine Japan's leadership in AI innovation, or that could inadvertently put Japan at odds with currently ongoing international standardization efforts. Japan can continue to position itself as a global leader in AI by adhering to its established policy objectives and

avoiding unnecessary or interventionist measures that are inconsistent with both its AI leadership goals and its existing regulatory or international standardization frameworks.

Risks of Using Code Compliance as a Public Procurement Criteria

As mentioned above, the draft Code, in (4) Other Matters, states that the Government is expected to evaluate the disclosures and specific initiatives of AI businesses and provide certain incentives in projects that the Government implements, suggesting compliance with the Code may impact public sector procurement. Given that the draft Code lacks a clear legal basis under Japanese law, doing so could create risks of inconsistency with market access obligations for goods and services under applicable international trade agreements, including Japan's commitments under the WTO Government Procurement Agreement, as well as corollary commitments under Japan's Economic Partnership Agreements and Regional Trade Agreements.

Encourage Responsible Use by AI Users

Generative AI systems are tools that can be used for a wide array of tasks. AI developers, as providers of a general-purpose tool, enable users of those tools to do a wide range of things. If specific users engage in conduct that rightsholders are concerned about, the appropriate response should focus on encouraging responsible and lawful use of AI tools by those users, rather than restricting the underlying technology itself — particularly, as mentioned earlier, where many AI models already incorporate safeguards such as output filtering and prompt input restrictions designed to reduce the risk of generating outputs that are substantially similar to existing copyrighted works. We encourage the Government to prioritize measures that promote appropriate use of AI systems.

Recommendation: The IP Strategy HQ should reaffirm its commitment to promoting the appropriate and responsible use of AI systems, including through use of technical safeguards, through enhanced user literacy and awareness, consistent with the guidance issued by the Agency for Cultural Affairs,[3] which states users' responsibility for outputs.

Conclusion

[Category of Comment (Relevant Section): (3) Others]

We thank the IP Strategy HQ for considering our comments. We strongly recommend IP Strategy HQ to continue discussion with affected stakeholders, including by providing the second public comment opportunity to respond to the amended draft Code which will be released after this public consultation. BSA looks forward to continued opportunity to work with the Government to advance AI innovation while ensuring appropriate protections for users, rightsholders, and other interests.

---

[3] General Understanding on AI and Copyright in Japan: https://www.bunka.go.jp/english/policy/copyright/pdf/94055801_01.pdf