Business Software Alliance's
Response to the European Commission Evaluation of the DSM Directive

June 2026

SURVEY #3 — Questions for Commercial Users of TDM

Section 2.1: Text and data mining (TDM), teaching exception, preservation exception

Use of TDM, educational and preservation exceptions (Articles 3–6).

For the next few questions please note that the Directive provides for an exception for text and data mining activities conducted for the purposes of scientific research (Article 3) and a general text and data mining exception (Article 4). Rightholders may reserve their rights and exclude their content from the scope of the latter exception.

Q4. Have you relied on the general TDM exception (Article 4) for your activities?

Instructions: Please select one option.

1. Yes
2. No
3. I don't know

Q5. To what extent does the general TDM exception meet your needs?

Instructions: Please select one option.

1. To a great extent
2. To some extent
3. To a little extent
4. Not at all
5. I don't know

Q5_1 [If Q5=1,2,3,4] Can you please elaborate on your response?

Most BSA members depend on the general TDM exception as a practical precondition for operating in today's data-driven economy. BSA's membership consists

predominantly of software application providers; companies that integrate AI capabilities into the business tools, productivity platforms, and sector-specific software that organisations across the economy rely on daily. Delivering those capabilities, and continuously improving them, requires the ability to analyse large volumes of diverse, lawfully accessed data without seeking permission for each individual work. That is not a peripheral legal question; it is the operational foundation on which AI-enabled software products are built.

The licensing alternative is not a realistic substitute, and it is important that the Commission understands why. BSA acknowledges that licensing arrangements exist for specific categories of high-value, actively managed content; paywalled databases, premium datasets, and curated specialist corpora; and those arrangements have their place. But they represent a small and structurally bounded segment of the data landscape. The universe of data required for modern TDM applications; and for training high-performing AI models in particular; is orders of magnitude larger than what is or ever will be available through negotiated licensing. The overwhelming majority of copyrighted works in existence are not actively managed by their rightholders. They are not listed in any catalogue, not offered for licensing, and in many cases, it is not practicable even to identify who holds the rights, let alone to contact them. This is true of legacy content, but it is equally true of the vast corpus of born-digital material; website content, user-generated text, forum discussions, code repositories; that forms the practical backbone of modern training datasets. A requirement to clear rights to this material would not redirect value to rightholders; it would simply render the data unusable.

The exception is therefore not a carve-out from a functioning market; it fills a gap that the market cannot fill. Its value flows to all parties: developers gain access to the data their work requires; rightholders benefit from the innovation and economic activity that flows from AI development; and the public gains from products and services that would not otherwise exist. Where commercial transactions are viable — for high-value, actively managed content where rightholders have a genuine interest in licensing and the market can support it — BSA supports the Commission encouraging market-led solutions alongside the exception. A narrow or uncertain interpretation of Article 4 does not protect any of these interests; it simply leaves the data, and its potential, untapped.

BSA members are themselves significant rightholders. The software industry depends on copyright to protect the code, documentation, datasets, and creative outputs that represent its core investment and competitive advantage. BSA therefore approaches this framework not as an actor seeking to circumvent intellectual property rights, but as one with a direct stake in a copyright system that functions well for everyone; creators, innovators, and the public alike. A well-calibrated TDM exception is not in tension with strong copyright protection; it is part of what makes the overall framework coherent and fit for a digital economy. It is precisely because BSA members understand the value of legal certainty in IP frameworks that they are so concerned by the uncertainty that continues to surround Article 4 in practice.

BSA members operate in a competitive global market where the United States, Japan, and Singapore have adopted broad TDM freedoms through fair use doctrines or statutory exceptions. If Article 4 is interpreted narrowly or burdened by unpredictable opt-outs, companies will site their AI training infrastructure and R&D investment outside the EU. The Commission should treat a harmonised, innovation-friendly interpretation of Article 4 not merely as a legal question but as a strategic industrial policy imperative; and one with particular urgency for SMEs and scale-ups, for whom legal uncertainty functions as a structural barrier to entry that larger incumbents can absorb but smaller operators cannot.

Q6. Have you relied on the TDM exceptions in activities related to the development of large language models / AI models?

Instructions: Please select one option.

1. Yes

2. No

3. I don't know

Q6_1 [If Q6=1] If yes, can you please indicate concrete cases and elaborate on the benefits / limitations of the TDM exceptions?

Most BSA members rely on TDM exceptions at every stage of AI development; not as a peripheral legal convenience, but as the operational foundation on which model development depends. Training foundational models requires analysing vast and diverse datasets drawn from publicly available and third-party sources, the majority of which are copyright protected. Fine-tuning those models for specific applications requires mining domain-specific corpora: scientific and medical literature, technical standards, regulatory filings, legal texts. Neither activity is feasible at the scale required if rights must be individually cleared.

The breadth of downstream reliance is worth underscoring. BSA members build software and AI infrastructure that powers applications across the entire economy; from fraud detection and credit risk modelling in financial services, to diagnostic support and drug discovery in healthcare, to predictive maintenance and quality assurance in manufacturing. In each case, the AI capability being deployed was built on TDM-enabled training. The exception does not just benefit AI developers; it is the upstream condition for AI-powered innovation in every sector that uses it.

A point that is sometimes lost in policy debate is why large-scale, diverse training data is technically necessary in the first place. The dominant insight driving AI progress over the past decade is that model performance scales with both computational resources and the breadth and diversity of training data. Domain-specific data alone does not produce robust, general-purpose models. The models that underpin the most valuable AI applications today were trained on tens of billions of works drawn from across the web, the scientific literature, and specialist corpora. Restricting access to that breadth of data does not produce a more targeted AI; it produces a weaker one.

Three use cases are of particular importance to BSA members and warrant explicit recognition in the Commission's evaluation:

- **AI-assisted software development.** Code generation, automated debugging, and vulnerability scanning tools are trained on large corpora of publicly available source code and technical documentation, much of which is copyright protected. These tools are foundational to developer productivity and directly relevant to the EU's Digital Decade objectives. If TDM exceptions do not reliably cover software artefacts, such tools cannot be legally built or improved within the EU.
- **Cybersecurity and threat intelligence.** Effective cybersecurity depends on the ability to mine malware samples, vulnerability disclosures, network logs, and threat intelligence feeds; much of which attracts copyright protection. Given the EU's strategic investment in cybersecurity under NIS2 and the EU Cybersecurity Act, the Commission should clarify that TDM conducted for cybersecurity research purposes falls within the Article 3 scientific research exception, which is not subject to rightholder opt-outs. Treating security research as a purely commercial TDM activity subject to Article 4 opt-outs would be both legally questionable and strategically damaging, given the public interest character of the work.
- **AI embedded in enterprise software.** BSA members provide the ERP systems, CRM platforms, cloud databases, and productivity tools used by businesses across the economy. These products increasingly incorporate AI capabilities whose development involves TDM at scale. The Commission should clarify that TDM for ongoing model improvement within enterprise software products falls within the scope of Article 4.

The global competitiveness dimension cannot be treated as a background consideration. The United States, Japan, and Singapore each provide materially broader TDM freedom than the EU's current framework delivers in practice. Companies make deliberate decisions about where to site model training infrastructure, and legal certainty - or its absence - is a significant input into those decisions. Fragmented opt-outs and interpretive divergence across Member States are not abstract compliance problems; they are factors that make the EU a less attractive location for AI R&D investment.

The practical limitations BSA members encounter reflect these structural problems directly: rights reservations that are technically undetectable; opt-out volumes that significantly reduce the usable share of otherwise lawfully accessible datasets; platform restrictions that impede access to content for which no reservation has been declared; and compliance standards that vary across Member States, creating overhead for any operator working at EU scale. These are not fringe concerns; they affect the day-to-day operation of AI development pipelines and disproportionately burden smaller companies that lack the legal and technical resources to manage them.

Q7. [If Q2=yes and/or Q4=yes] Have you faced any barriers in making use of the TDM exceptions?

Instructions: Please select all applicable options.

1. I have not faced any barriers
2. Legal uncertainty about my use case
3. Difficulties to obtain lawful access to the works
4. Technical restrictions and Technological Protection Measures (TPMs) preventing TDM on content to which I have lawful access
5. TDM not allowed under licence or contract
6. Difficulties to identify rights reservations expressed by rightholders
7. High volume of content for which rights were reserved ('opted out')
8. Other, please specify

Q8. How has Article 4 impacted your TDM practices and / or business model?

Instructions: Please select one option.

1. Negatively
2. Somewhat negatively
3. Positively
4. Somewhat positively
5. It hasn't impacted them at all
6. I don't know

Q9. Please elaborate on your assessment and explain for which type of activities you have relied on the TDM exception (e.g. data analytics, AI training, etc) and in which sector.

The TDM exception is not relied upon for a single category of activity. For BSA members it operates as a cross-cutting legal foundation; present at every stage of the AI development lifecycle and embedded in the products and services that carry AI capability into the broader economy.

At the model development layer, this means pre-training on large and diverse datasets, fine-tuning on domain-specific corpora, continuous evaluation against new data, and ongoing improvement of deployed models as real-world usage surfaces gaps and errors. Each of these stages requires the ability to analyse substantial volumes of lawfully accessed, copyright-protected content without individually negotiating permissions; a requirement that is not a design choice, but a technical and economic necessity given the scale at which modern AI operates.

What is less well understood in policy discussions is the position BSA members occupy in the AI value chain. BSA's membership consists predominantly of software application providers; companies that integrate AI capabilities into the business tools, productivity platforms, and sector-specific software that organisations across the economy rely on daily. They are neither foundation model developers at one extreme nor passive end-users at the other, but the layer through which AI reaches the vast

majority of its actual users. The sectors that appear most prominently in assessments of AI adoption; healthcare, financial services, manufacturing, energy; do not build their own AI from scratch; they deploy it through the business software applications that BSA members build and maintain. A restriction on TDM therefore does not land only on AI developers: it propagates downstream to every organisation that depends on AI-enabled software to operate. At a time when the EU is looking to accelerate Applied AI adoption, and when the uptake in certain sectors would support the achievement of other EU objectives (i.e. energy efficiency in housing), such a restriction would frontally contradict this roadmap.

This layered structure also has implications for how compliance burdens are distributed. Large software companies building foundation models have legal and technical teams capable of managing complex rights landscapes. The healthcare provider deploying an AI diagnostic tool, or the manufacturing SME using an AI-assisted quality control system, does not. The compliance overhead that Article 4's current implementation generates; through legal uncertainty about scope, technically undetectable opt-outs, and divergent national standards; falls ultimately on the entire chain, not just those at the top of it. Measures that reduce that overhead, including a clear interpretive standard on what constitutes a legally sufficient rights reservation, are therefore both a competitiveness measure and a question of who gets to benefit from AI at all.

The single most impactful step the Commission could take is to issue an interpretive communication confirming three things: that AI training constitutes TDM within the meaning of Article 4; that lawful access extends to lawfully published publicly available content; and that a rights reservation not expressed in a technically discoverable, machine-readable form at the time of access is not enforceable against a good-faith practitioner. That clarification would reduce compliance overhead across the entire chain, support consistent application across Member States, and send a clear signal that the EU intends to remain a competitive jurisdiction for AI development.

Q16. What unexpected benefits or challenges have you experienced in relation to the introduction of the TDM exceptions?

The most significant unexpected challenge has been the opt-out ecosystem. When the Directive was adopted, the Article 4 rights reservation mechanism was understood as a limited safeguard for rightholders with active, managed content and a genuine commercial interest in restricting TDM. In practice it has evolved into something considerably more disruptive: a fragmented, technically inconsistent landscape of blanket restrictions, many of which are expressed in forms that are legally ambiguous, technically undetectable, or applied retroactively to activities that were already underway. The result is that a significant portion of publicly accessible content; content that rightholders have made no active effort to license or commercialise; is nominally opted out in ways that create legal exposure without serving any discernible rightholder interest.

For BSA members, the practical consequences are concrete. Data pipelines built on content that carried no restriction at the time of access are subsequently exposed to claims based on reservations introduced after the fact. Compliance teams must monitor a constantly shifting landscape of platform-level restrictions, contractual terms, and technical signals that are neither standardised nor consistently interpretable. Smaller members; and the SME customers who depend on their AI-enabled products; bear these costs without the legal infrastructure to manage them effectively.

The solution does not require new EU legislation or a bespoke EU standard. The Robots Exclusion Protocol (robots.txt) is, today, the only mechanism that is both technically mature and widely recognised as a baseline for expressing machine-readable content restrictions; it is already implemented across the web, understood by developers, and operationally deployable today. What is missing is legal recognition. BSA urges the Commission to clarify that a rights reservation expressed through robots.txt constitutes a valid and legally sufficient reservation under Article 4, while signalling openness to recognising more granular successor mechanisms; and that a reservation not expressed in a technically discoverable form at the time of access cannot be enforced against a good-faith TDM practitioner. Building on something that already exists and is widely recognised is faster, more credible, and less susceptible to regulatory capture than designing something new.

BSA acknowledges that robots.txt was not designed with AI training in mind and does not on its own provide the granularity that rightholders or practitioners may ultimately want. Industry is actively engaged in developing more sophisticated and interoperable controls; including through ongoing multistakeholder work at the IETF on AI preferences; and BSA is committed to contributing to that process. The Commission should support rather than pre-empt that work: supporting robots.txt now, while recognising more granular successor mechanisms as they achieve genuine consensus and adoption, is the approach most likely to produce a stable and workable outcome.

That clarification should also establish what a valid reservation can and cannot do. It should operate prospectively; data lawfully accessed before a reservation was in place cannot be retroactively encumbered. It should be capable of granular expression by content type or use case, rather than functioning only as an all-or-nothing exclusion. And it should have no application to publicly funded research outputs, open-licensed content, or government data, which should remain within the exception regardless of any purported reservation.

The unexpected benefit, by contrast, has been the Directive's effect as a signal of intent. Enacting a general TDM exception was a meaningful commitment: it told the market that the EU intended to be a viable jurisdiction for data-driven innovation, and it gave European-based operators a legal foundation on which to build. That signal has had real value; but it is contingent on implementation delivering what the text promises. The risk BSA observes is not that the Directive was misconceived, but that interpretive divergence, opt-out fragmentation, and regulatory uncertainty are quietly eroding its

practical effect. The Commission's evaluation is an opportunity to address that erosion before it becomes irreversible.

Section 2.5: Other impacts

Q126. Based on your experience, please indicate one key benefit or value added by the DSM Directive, if any?

The most significant benefit of the Directive is one that tends to be underappreciated precisely because it operates as a foundation rather than a feature: the general TDM exception created a legal basis where none existed before, and in doing so made an entire category of economically valuable activity possible at EU scale. Prior to the Directive, the legal status of large-scale data analysis for commercial purposes was genuinely uncertain across much of the EU. That uncertainty had a chilling effect; on investment decisions, on product development, on where companies chose to locate AI-related operations. The exception did not resolve every question, but it established the principle, and that matters.

For BSA members specifically, the value is both direct and structural. Directly, it provides the legal foundation on which AI-enabled software products are built and continuously improved. Structurally, it positions BSA members; as software application providers whose products carry AI capability into every sector of the economy; to serve customers in healthcare, financial services, manufacturing, and beyond, without those customers needing to navigate copyright exposure themselves. The exception's benefits compound through the value chain.

The competitive parity dimension also deserves recognition. The United States had long permitted broad TDM activity through fair use; Japan introduced a specific TDM exception in 2019; Singapore followed a similar path. The DSM Directive brought the EU meaningfully closer to that baseline, reducing the legal asymmetry that had previously made European jurisdictions less attractive for AI R&D investment. BSA members with global operations experienced this shift directly; activities that had previously required careful legal structuring to locate outside the EU became viable within it.

BSA's concern is that this progress is more fragile than it appears. The competitive parity the Directive achieved is not self-sustaining: it depends on consistent, broad implementation across Member States, and on the exception’s scope being interpreted in a manner that keeps pace with how AI development actually works. Divergent national implementation, aggressive opt-out practices, and narrow judicial interpretations of "lawful access" or "machine learning" each erode the single market advantage the Directive was designed to create. The Commission should treat the preservation of that advantage as an active responsibility; monitoring implementation outcomes and being prepared to act through interpretive guidance or, where necessary, infringement proceedings, if the exception's practical scope is being quietly narrowed at the national level.

Q127. To what extent has your organisation incurred compliance costs because of the Directive?

Instructions: Please select one option.

1. To a great extent
2. To some extent
3. To a little extent
4. Not at all
5. I don't know

Q128. What is the nature and level of those compliance costs and in relation to which of the Directive's provisions are they incurred?

The compliance costs associated with Article 4 are more substantial and more unevenly distributed than a reading of the Directive's text would suggest. The theoretical framework; analyse lawfully accessed content, respect declared opt-outs, retain documentation; is straightforward. The operational reality is considerably more complex. Determining what constitutes lawful access in a given context, identifying whether a rights reservation has been declared and in what form, assessing whether a licence's silence on TDM amounts to a prohibition, and adapting data pipelines when the legal picture shifts; each of these requires legal judgment, technical capability, and ongoing monitoring. For companies operating at scale across multiple content sources and multiple Member States, the aggregate burden is significant.

The distribution of that burden is where the structural problem lies. Large operators can absorb it: they have legal teams, compliance infrastructure, and the technical capacity to build rights-detection into their data pipelines. Smaller companies; including the SME software providers that make up a meaningful share of BSA's membership, and the SME customers who deploy their AI-enabled products; generally cannot. The result is that compliance overhead functions as a barrier to entry, concentrating TDM activity among incumbents and penalising precisely the companies that a competitive, innovative AI ecosystem most needs to thrive.

Two measures would materially address this. The most impactful by some distance would be a Commission clarification that rights reservations must be expressed in a technically discoverable, machine-readable form - such as robots.txt - to be enforceable against a good-faith TDM practitioner. This would not eliminate compliance costs, but it would make them manageable and predictable: operators would know what to look for, where to look, and what the legal consequence of not finding a reservation is. The current landscape, in which reservations may be buried in general terms of service, asserted retroactively, or expressed in forms that automated systems cannot detect, is the primary driver of the disproportionate costs that smaller operators face.

The second measure is a safe harbour for content accessed in good faith before a machine-readable reservation was in place. Retroactive exposure is one of the most

significant sources of legal uncertainty BSA members report: the risk that a data pipeline built on content that carried no detectable restriction at the time of access is subsequently challenged on the basis of a later or previously undetectable reservation. A clear prospectivity rule would eliminate that exposure.

Q129. How well do the DSM Directive's provisions align with other relevant EU legislation (e.g., GDPR, AI Act, Digital Services Act, Orphan Works Directive)?

Instructions: with 1 being not at all aligned and 5 being fully aligned.

1. Slider: 1 **2** 3 4 5

Q129_1 [If Q129=1 or 2] Please specify any particular laws where you have noticed alignment or conflict.

The DSM Directive was adopted before the AI Act was conceived, and the gap shows. The two instruments now govern overlapping activities; the training of AI models on copyright-protected content; through frameworks that were developed independently and have never been reconciled. The result is a set of tensions that are not merely theoretical: they create genuine compliance dilemmas for BSA members who must simultaneously satisfy obligations under both regimes.

The most acute point of friction is the AI Act's requirement that providers of general-purpose AI models document and disclose their training data. BSA does not oppose transparency as a policy objective. But the interaction between that obligation and Article 4 has not been thought through, and three specific problems require the Commission's attention.

The first is the liability question. A company that discloses its training data sources in compliance with the AI Act should not, by virtue of that disclosure, be taken to have acknowledged that those sources were used in breach of an Article 4 opt-out or otherwise in violation of copyright. The two obligations serve entirely different purposes; one is about market transparency and AI accountability, the other is about intellectual property; and compliance with one should not generate exposure under the other. Without an explicit clarification to this effect, disclosure becomes a legal risk rather than a neutral regulatory requirement, which will predictably lead to over-compliance and reduced transparency, the opposite of what the AI Act intends.

The second is definitional consistency. The AI Act refers to "training data" and the DSM Directive refers to "works or other subject matter." These are not defined in a coordinated way, and there is a real risk of definitional drift; where the same underlying activity is characterised differently under each instrument, creating compliance gaps or conflicting obligations for operators trying to satisfy both simultaneously. The Commission should ensure that guidance under both frameworks uses consistent terminology and scope.

The third concerns retention. The AI Act requires providers to maintain documentation of training data for accountability and audit purposes. Article 4(3) of the DSM Directive limits the retention of copies of works to what is necessary for the TDM activity. These obligations point in different directions, and without a carve-out, operators face an irresolvable conflict: retain documentation to satisfy the AI Act, or delete it to comply with the DSM Directive. This is precisely the kind of incoherence that discredits regulatory frameworks and imposes costs with no corresponding benefit to anyone.

BSA recommends that the Commission convene an inter-institutional working group to address these conflicts and produce operationally workable guidance within twelve months of the AI Act's full application date. The objective should not be to subordinate one framework to the other, but to map the points of interaction clearly enough that companies can comply with both without having to choose between them.

Q130. Have you experienced any legal or practical tensions or conflicts between the DSM Directive as implemented and other national laws in your country?

Instructions: Please select one option.

1. Yes

2. No

3. I don’t know

Q130_1 [If Q130=Yes] Please describe the tension or conflict and its impact.

The DSM Directive is a harmonisation instrument, but its effect in practice has been anything but harmonised. BSA members operating across multiple EU jurisdictions encounter materially different legal landscapes depending on where they are. The concepts that matter most ; what constitutes "lawful access," whether AI training falls within the scope of "text and data mining," what form a rights reservation must take to be valid ; have been interpreted differently by national courts and regulatory bodies across Member States, in some cases in ways that are considerably narrower than the Directive's text warrants. For a company running a single data pipeline that ingests content from across the EU, this divergence does not produce legal nuance; it produces unmanageable exposure.

The tensions with other regulatory frameworks compound this problem. BSA members increasingly find themselves caught between obligations that pull in opposite directions: content that must be processed to meet accessibility or safety requirements under one instrument, but that a copyright opt-out purports to restrict under another; documentation that must be retained for AI Act compliance purposes, but that copyright rules suggest should be deleted. These are not edge cases. They are foreseeable consequences of a legislative landscape that has grown faster than the mechanisms for ensuring its internal consistency.

The minimum harmonisation approach that characterises the current Directive has run its course for TDM. The variation it permits may have been tolerable when TDM was a

specialist research activity; it is not tolerable when TDM underpins AI development pipelines that operate at EU scale and cannot be practically restructured jurisdiction by jurisdiction. BSA urges the Commission to assess whether maximum harmonisation of the Article 4 exception; a single, directly applicable standard that leaves no room for the national interpretive divergence that currently distorts the single market; is the appropriate legislative response.

In the interim, two steps would materially improve the situation. The first is an authoritative interpretive communication from the Commission clarifying the scope of Article 4: confirming that AI training constitutes TDM, that lawful access extends to publicly available content, and that only machine-readable rights reservations that are technically discoverable at the time of access are enforceable against good-faith practitioners. The second is the Commission's willingness to use infringement proceedings where national implementation is demonstrably inconsistent with the Directive's objectives. The single market in data cannot function if Member States are effectively permitted to opt out of the harmonisation the Directive was designed to deliver.

* * *

For more information, please contact:
Thomas Boué: thomas@bsa.org