



Trends and Developments

Contributed by:

Miguel Geijo and Antonio Borjas
Broseta Abogados

Broseta Abogados is a full-service law firm advising leading companies and institutions across a broad range of practice areas and sectors, with particular expertise in financial services, energy, life sciences, real estate and the digital economy. Its Privacy, IT and Digital Environments practice comprises 14 lawyers, led by partners Miguel Geijo and Agustín Puente, primarily based in Madrid, Barcelona and Valencia. The team advises listed companies, financial institutions and multinational groups on complex data protection matters, including GDPR and LOP-DGDD compliance, privacy governance, international

data transfers, security breaches, biometric data, digital advertising and data-driven business models, with particular expertise in artificial intelligence and emerging technologies. Recent work includes advising Banco Santander, Openbank, BBVA, CaixaBank, Iberdrola, Samsung and Novo Nordisk on compliance and strategic privacy projects. The team also advises leading industry associations, including AEB, Farmaindustria, UNESPA and AELEC, and represents clients before the Spanish Data Protection Authority and courts.

Authors



Miguel Geijo is a partner at Broseta and head of the firm's Commercial Law department. He specialises in information technology law, data protection and privacy, with extensive experience advising on the legal

implications of artificial intelligence, emerging technologies and digital transformation. He is a former Secretary of the Board of the Spanish Association of Privacy Professionals (AEPD). He also lectures on privacy, intellectual property and new technologies, including on the Master's Degree in Intellectual Property, Industrial Property and New Technologies at the Autonomous University of Madrid, as well as on other postgraduate programmes.



Antonio Borjas is a senior associate in Broseta's Commercial Law and Privacy, IT and Digital Environments practices, specialising in information technology, data protection, privacy and digital regulation. He advises

clients on legal and regulatory matters arising from the development and use of new technologies, including artificial intelligence, as well as on compliance and information security matters. He is a certified Data Protection Officer under the Spanish Data Protection Agency's (AEPD) certification scheme, holds the CESCO Compliance Certification and a specialisation in Compliance from IEB, and has lectured on data protection for the Spanish Association of Privacy Professionals (AEPD).

Broseta Abogados

Paseo de la Habana 101,
28036, Madrid
Spain

Tel: +34 91 432 3144
Email: info@broseta.com
Web: www.broseta.com



Traceability and Technical Documentation as the Foundation for Copyright Protection in Artificial Intelligence Systems

A shift in focus: from ownership of rights to proof of facts

The debate on copyright and artificial intelligence (AI) has evolved. The focus is increasingly shifting from abstract questions about whether protected works can be used to train AI models to a more practical issue: who can prove what, and when.

The training of generative and general-purpose AI (GPAI) models relies on vast volumes of content, much of which may be protected by copyright. As a result, the debate is no longer limited to the theoretical scope of copyright protection. It also concerns the practical ability to reconstruct, sometimes years later, which data were used, under what licence or legal basis, and with what result. In this context, the party able to generate and preserve reliable technical evidence over time may have a significant legal advantage.

Regulation (EU) 2024/1689 (“AI Act”) is not a copyright statute and does not create new intellectual property rights. Its contribution to copyright protection is indirect but potentially significant. The obligations it imposes in areas such as data governance, technical documentation, record-keeping and transparency may generate the evidence needed to identify data sources, reconstruct how an AI system operates and establish relevant facts in litigation.

Traceability, therefore, can serve a dual purpose. Although conceived under the AI Act primarily as a tool for safety, compliance and risk management, it may also function as an evidentiary infrastructure supporting copyright protection.

This article examines how this evidentiary architecture operates from the perspective of Spanish law, although the underlying logic may also be relevant in other jurisdictions. A brief terminological clarification may be useful for international readers. Under Spanish law, “intellectual property” (*propiedad intelectual*) refers to copyright and related rights governed by the Spanish Intellectual Property Act (TRLPI), while patents, trade marks and designs fall within the separate

category of “industrial property”. This article concerns the former.

The AI Act as an evidentiary infrastructure

When analysing the AI Act from a copyright perspective, it is important not to conflate different regulatory regimes. Obligations concerning data governance under Article 10, technical documentation under Article 11 and Annex IV, automatic event logging under Article 12, and transparency towards deployers under Article 13 apply to high-risk AI systems. They do not apply to all AI systems or to every generative model.

This distinction is essential. Otherwise, there is a risk of overstating the traceability obligations imposed by the AI Act and misunderstanding how different categories of AI systems may generate evidence relevant to copyright disputes.

In practice, copyright-relevant traceability may arise through different regulatory layers. For high-risk AI systems, the connection with copyright is mainly indirect and evidentiary. Requirements concerning data governance, technical documentation and logging are not designed specifically to protect copyright. Nevertheless, the information generated through compliance with those requirements may later become relevant in a copyright dispute.

For GPAI models, the connection is more direct. Article 53 expressly addresses compliance with Union copyright law and transparency regarding training content. A further layer is provided by Article 50, which establishes transparency obligations for certain AI systems regardless of whether they are classified as high-risk.

Nonetheless, as indicated above, not every AI system falls within this evidentiary architecture. Many systems will neither qualify as high-risk, fall within Article 50, nor be provided as GPAI models. For those systems, the AI Act does not impose the same documentation, logging or transparency requirements that could later support copyright enforcement.

In such cases, evidence generation will depend primarily on general copyright and text and data mining (TDM) rules, together with voluntary governance practices and contractual arrangements adopted by

providers and rightsholders. Information asymmetry may therefore remain considerably higher, and the ability to reconstruct relevant facts after the event will depend largely on the parties' own strategic choices.

Aside from the cases mentioned above, for high-risk systems, Article 10 of the AI Act requires data governance practices concerning the provenance, collection and processing of training, validation and testing datasets. If those datasets contain protected works or materials subject to contractual restrictions, this documentation may become an important first link in the evidentiary chain.

Article 11 of the AI Act, together with Annex IV, requires technical documentation covering aspects of the system's design, development and evaluation. This information may help reconstruct relevant technical processes and identify the metrics used. Its importance for copyright is therefore contingent rather than inherent: these provisions are not copyright rules, but they may preserve information capable of establishing facts relevant to a later copyright claim.

Article 12 of the AI Act adds another layer by requiring automatic logging for high-risk systems. These logs create a traceable record of system operation and are primarily intended to support post-market monitoring and regulatory supervision. However, they may also acquire evidentiary value when a copyright dispute concerns the operation of a system rather than, or in addition to, its training.

Retention periods are particularly relevant in this context. Under Article 18 of the AI Act, providers must retain technical documentation for ten years after the system has been placed on the market or put into service. Automatically generated logs must also be retained for a period appropriate to the system's intended purpose and in accordance with the applicable requirements under the AI Act and relevant sectoral legislation, including the obligations imposed on providers under Article 19 and deployers under Article 26 (6).

The ten-year retention period for technical documentation is particularly significant from a copyright enforcement perspective in Spain. Although copy-

right economic rights may subsist for the life of the author plus 70 years, civil actions for damages arising from copyright infringement are generally subject to a five-year limitation period under the general rules on personal actions (Article 140 (3) TRLPI). In practical terms, relevant technical documentation will therefore typically remain available beyond the limitation period applicable to most damages claims.

The AI Act's transparency framework is not limited to high-risk systems. Article 50 of the AI Act establishes obligations for certain AI systems regardless of their risk classification. For example, providers of systems intended to interact directly with individuals, such as chatbots, must ensure that users are informed that they are interacting with an AI system. Providers of systems generating synthetic audio, image, video or text content must also ensure that outputs are appropriately marked as artificially generated or manipulated.

These requirements are not specifically designed to protect copyright. Nevertheless, the identification and labelling of AI-generated or AI-manipulated content may have an evidentiary function. They may help distinguish synthetic or manipulated material from other content and, in appropriate cases, establish factual circumstances relevant to disputes involving authorship, reproduction or the use of protected works. They do not, however, determine whether copyright infringement has occurred.

GPAI and copyright: the role of TDM

The AI Act provisions most directly concerned with copyright are those applicable to GPAI models, which have applied since 2 August 2025.

First, Article 53 (1)(d) of the AI Act requires providers to prepare and make publicly available a sufficiently detailed summary of the content used to train their models, following a template provided by the AI Office. These obligations do not create new rights for rightsholders. They may, however, reduce information asymmetries and facilitate the enforcement of existing rights.

On 24 July 2025, the European Commission published the Explanatory Notice and Template for the Public

Summary of Training Content, establishing a mandatory common baseline for the information that GPAI model providers must disclose. Developed following a multi-stakeholder consultation, the template is intended to increase transparency regarding training data and to assist parties with legitimate interests, including copyright holders, in exercising their rights under Union law.

Alongside this framework, the General-Purpose AI Code of Practice, published on 10 July 2025, provides a voluntary benchmark that providers may use to demonstrate compliance with the AI Act's GPAI obligations concerning transparency, copyright and safety.

Second, Article 53 (1)(c) of the AI Act requires providers to put in place a policy to comply with Union copyright law. Crucially, this includes identifying and respecting the reservation of rights provided for in Article 4 (3) of Directive (EU) 2019/790 on copyright and related rights in the Digital Single Market ("DSM Directive").

The reference to rights reservations in Article 53 (1) (c) of the AI Act brings a central element of the EU copyright framework into the GPAI regime: the TDM exceptions.

Articles 3 and 4 of the DSM Directive establish exceptions for TDM. Article 4 permits reproductions and extractions of lawfully accessible works and other subject matter for TDM purposes, provided that the use has not been expressly reserved by rightsholders in an appropriate manner. In Spain, these provisions were transposed through Royal Decree-Law 24/2021.

Where a provider seeks to rely on the Article 4 TDM exception in connection with AI training, its applicability may therefore depend on factual questions. Was there a valid reservation of rights? Was that reservation expressed in an appropriate manner? Could the provider identify it? And was it respected during the collection and processing of training data?

It is here that traceability acquires its clearest copyright-specific evidentiary function. A legal framework that allows rightsholders to reserve their rights has

limited practical effectiveness if the relevant facts cannot subsequently be reconstructed. This may require evidence concerning the source of training content, the existence and form of any rights reservation, and the provider's response to that reservation.

The decisive issue, therefore, is not only how rights and obligations are defined in law, but whether the facts necessary to establish compliance, or infringement, can be proven.

AESIA and the translation of regulation into verifiable practices

The Spanish Agency for the Supervision of Artificial Intelligence (AESIA) is the public body responsible for AI supervision in Spain. Among its activities, AESIA has published a set of 16 practical guidelines aimed at facilitating the implementation of the AI Act's requirements.

Their relevance to copyright does not lie in creating new rights or copyright-specific obligations. Rather, they may help translate regulatory requirements on documentation, data governance, record-keeping and transparency into concrete traceability practices.

Among AESIA's 16 guidelines, those concerning data governance, technical documentation, record-keeping and logging, transparency and quality management systems are particularly relevant from an evidentiary perspective.

These practices may help create and preserve information capable of reconstructing important aspects of an AI system's lifecycle. Documentation on datasets and their provenance may help establish what data were used and where they came from. Records and logs may provide evidence of relevant technical processes and decisions. Technical documentation and version control may also help identify the datasets, system components and model versions involved at a particular point in time.

From a copyright perspective, this information may be particularly relevant when determining whether protected works were used in the development or training of an AI system. Where applicable, it may also help

establish whether reservations of rights under Article 4 of the DSM Directive were identified and respected.

Traceability may also assist in reconstructing the circumstances surrounding the generation of particular outputs and in identifying the actors and technical processes involved.

AESIA's guidelines do not themselves determine whether copyright infringement has occurred. Nor are they rules of copyright law or, by themselves, evidence of infringement. Their significance lies elsewhere: they illustrate how AI governance and compliance practices may generate and preserve information that can later support the investigation and proof of copyright-related claims.

In this sense, the traceability practices reflected in AESIA's guidelines may contribute to an evidentiary infrastructure that supports the effective enforcement of copyright and related rights in the context of AI systems.

The dual function of traceability

Traceability can protect both rightsholders and AI providers.

For rightsholders, the ability to identify data sources, document datasets and verify compliance with the Article 4 DSM Directive rights reservation may be decisive when assessing whether the use of a protected work was lawful. The GPAI obligations concerning training content summaries and copyright policies are particularly relevant because they seek to reduce the information asymmetry that has historically made it extremely difficult for rightsholders to determine whether their works were used in AI training.

Traceability is also relevant to generated outputs. Adequate records may help reconstruct the circumstances surrounding a particular output, identify the model version involved and analyse the relationship between inputs and results.

This can be important when allegations concern substantial reproduction of a protected work, memorisation of training content or outputs that closely resemble pre-existing materials. Without adequate

traceability, rightsholders may face significant difficulties in meeting the burden of proof. With it, they may have a more objective basis on which to build a claim, request information or negotiate a licence.

Liability for AI-generated outputs nevertheless remains subject to the general legal framework. The proposal for an AI Liability Directive was withdrawn in February 2025. AI-related harm is now addressed, in part, through the revised Product Liability Directive (Directive (EU) 2024/2853), adopted on 23 October 2024, but there is no self-standing AI liability regime.

Responsibility for copyright-infringing outputs must therefore be assessed under existing contractual and tort rules, together with any other applicable legal frameworks, rather than under a dedicated AI liability statute.

Traceability also benefits AI providers. Technical documentation and instructions for use define a system's intended purposes, capabilities and conditions of use. This may help providers demonstrate contractual breaches, unauthorised uses or third-party manipulation.

Operational records may also provide relevant evidence in cases involving mass extraction of outputs, abusive use of interfaces, reverse-engineering attempts or circumvention of access restrictions. In this way, traceability may strengthen a provider's position when responsibility needs to be allocated among different actors.

Importantly, regulatory compliance does not necessarily require providers to disclose their most sensitive strategic assets. The information required for traceability will not normally involve public disclosure of model weights, internal architectures or detailed proprietary training processes.

Spain's Trade Secrets Act and Article 78 of the AI Act, concerning confidentiality, provide mechanisms for protecting sensitive information. In principle, therefore, providers can comply with documentation and transparency obligations while preserving the confidentiality of their core know-how.

Practical recommendations and limitations

The effectiveness of this evidentiary architecture ultimately depends on concrete organisational and technical measures.

In practice, providers should consider:

- Maintaining up-to-date inventories of data sources and licences, including the legal basis for each dataset used in training.
- Systematically versioning training datasets so that the composition of each training run can later be reconstructed.
- Documenting development processes, including key design decisions, evaluation metrics, model iterations and relevant third-party components.
- Preserving records with appropriate integrity safeguards, such as timestamping, digital fingerprinting and chain-of-custody procedures, to strengthen their evidentiary value in potential proceedings.
- Implementing the copyright policy required by Article 53 (1)(c) of the AI Act, including effective mechanisms for identifying and respecting opt-out reservations under Article 4 (3) of the DSM Directive.
- Preparing and publishing the training content summary using the mandatory template published by the AI Office on 24 July 2025, while ensuring consistency between internal documentation and public disclosures.

Providers may also use automated data-tracking tools and controlled data pipelines to support these measures. Such tools can help identify opt-outs and track how content moves through training and fine-tuning processes.

Rightsholders can also take proactive steps. In particular, they should consider:

- Exercising the Article 4 DSM Directive opt-out through appropriate means, including machine-readable reservations for content made publicly available online.
- Monitoring GPAI providers' published training content summaries to identify potential uses of their works and inform enforcement or licensing strategies.

- Preserving evidence of their rights, including registration records, publication dates and licence terms, to support future claims.

Traceability nevertheless has limits.

First, the principle of proportionality prevents record generation from becoming an indiscriminate collection of information without a legitimate purpose. Where traceability mechanisms involve personal data, data protection law must also be considered.

In particular, extensive logging may come into tension with the principles of data minimisation and storage limitation under Article 5 of the General Data Protection Regulation (GDPR). Providers must therefore identify an appropriate legal basis under Article 6 of the GDPR and comply with any other applicable data protection requirements.

Second, traceability must be balanced against the protection of trade secrets and confidential business information. This may require controlled-access mechanisms, restricted disclosure, confidential expert examination and procedural safeguards. The objective should be to allow relevant facts to be verified without unnecessarily exposing sensitive assets.

Looking ahead: an emerging trend

Effective copyright protection in the age of AI increasingly depends not only on how rights are defined, but also on whether reliable technical evidence can be generated and preserved.

The AI Act, together with practical guidance such as that published by AESIA, may contribute to an evidentiary infrastructure originally designed primarily for regulatory compliance, safety and risk management. Data governance, technical documentation, record-keeping and transparency may also provide the tools needed to reconstruct relevant aspects of an AI system's lifecycle, identify data sources, analyse outputs and establish the factual basis for allocating responsibility.

The emerging framework for GPAI models reinforces this trend. The copyright policies required under Article 53 of the AI Act, the General-Purpose AI Code

of Practice and the mandatory template for training content summaries contribute to a more structured transparency framework.

Together, these instruments may provide rightsholders with information that helps them assess whether their content may have been used in training and exercise their rights under Union law. At the same time, they may help providers demonstrate compliance and preserve evidence relevant to their own legal position.

The coming years will test how effective this architecture is in practice. Regulatory enforcement, litigation involving training content summaries, and emerging case law on the scope and effectiveness of TDM opt-outs will help determine the true strength of this evidentiary framework.

For rightsholders, greater traceability may provide clearer avenues for investigating and substantiating potential infringements. For providers, it may offer a means of demonstrating compliance and preserving evidence relevant to the allocation of responsibility.

In the digital economy, traceability may therefore become more than a regulatory compliance requirement. It may increasingly serve as a condition for legal certainty and as a strategic asset for all parties involved.