BRIEFING

Requested by the JURI Committee

# Technological Aspects of Generative AI in the Context of Copyright

## Attribution and Novelty in Generative AI Hypersurfaces

Policy Department for Justice, Civil Liberties and Institutional Affairs
Directorate-General for Citizens' Rights, Justice and Institutional Affairs PE 776.529 - July 2025

EN

# Technological Aspects of Generative AI in the Context of Copyright

## Attribution and Novelty in Generative AI Hypersurfaces

### Abstract

This in-depth analysis explains the statistical nature of generative AI and how training on copyright-protected data results in persistent functional dependencies with respect to the used data. It highlights the challenges of attribution and novelty detection in these high-dimensional models, emphasising the limitations of current methodologies. The study provides technical recommendations for traceability and output assessment mechanisms.

This study is commissioned by the European Parliament's Policy Department for Justice, Civil Liberties and Institutional Affairs at the request of the Committee on Legal Affairs.

This document was requested by the European Parliament's Committee on Legal Affairs.

## AUTHOR

Dr Axel BRANDO, HPES Lab, Barcelona Supercomputing Center

## ADMINISTRATOR RESPONSIBLE

Mariusz MACIEJEWSKI

## EDITORIAL ASSISTANT

Ivona KLECAN

## LINGUISTIC VERSIONS

Original: EN

## ABOUT THE EDITOR

Policy departments provide in-house and external expertise to support EP committees and other parliamentary bodies in shaping legislation and exercising democratic scrutiny over EU internal policies.

To contact the Policy Department or to subscribe for updates, please write to:
Policy Department for Justice, Civil Liberties and Institutional Affairs
European Parliament
B-1047 Brussels
Email: poldep-iust-b@europarl.europa.eu

Manuscript completed in July 2025

## DISCLAIMER AND COPYRIGHT

# CONTENTS

# LIST OF ABBREVIATIONS

| | |
|---|---|
| AI | Artificial Intelligence |
| AI ACT | Regulation (EU) 2024/1689 Artificial Intelligence Act |
| TDM | Text and Data Mining |
| DSMD | Digital Single Market Directive |
| GenAI | Generative Artificial Intelligence |

# LIST OF FIGURES

# EXECUTIVE SUMMARY

Generative Artificial Intelligence (GenAI) systems represent a fundamental shift in how digital content is produced. Unlike traditional predictive systems, which aim to infer properties or features of an input (typically without engaging with copyright-protected material), GenAI models are designed to generate human-like content across domains such as text, imagery, and music; domains in which intellectual property rights are often implicated.

Technically, a GenAI model can be understood as a high-dimensional hypersurface, a deformable grid within the space shaped by the training data points. During training, this grid is progressively deformed to approximate the points in the dataset, ideally passing close to - or, in some cases, exactly through – them, without forcing sharp or irregular bends. This smooth deformation enables the model to generalize beyond the training data. Once trained, generating content means evaluating this hypersurface at particular coordinates. This sampling process may yield outputs that appear entirely novel or may produce outputs that are indistinguishable from or identical to original training data.

The core technical challenge lies in the functional dependency that exists between the training data and the learned hypersurface at the time of generation. Since generating an output corresponds to selecting a point on the surface, and since that surface was itself shaped by the training data, the relationship between any output and the original data that influenced it is not directly traceable but probabilistically persists within the model's generative function. Each generated point is a consequence of the cumulative statistical influence of the training data that deformed the surface.

Two key copyright-relevant issues arise from this dependency:

1. Attribution: How can we determine whether a generated output is meaningfully derived, either partially or substantially, from one or more training examples? Current models offer no reliable mechanism to quantify or trace this influence.
2. Novelty: How can we assess whether an output constitutes a genuinely new creation, or if it is the result of high-dimensional stochastic parroting, i.e. statistically reproducing patterns with near-verbatim similarity under the guise of novelty?

This study invites a fundamental rethinking of how influence, derivation, and originality are interpreted in the era of GenAI. It highlights the need to re-evaluate the role and scope of legal frameworks such as the Text and Data Mining (TDM) exception introduced in the Digital Single Market Directive (Directive (EU) 2019/790), especially when models trained under such exceptions produce outputs that may raise questions of copyright infringement.

Addressing these challenges is not merely a matter of legal interpretation; it requires direct technical engagement. Ensuring the accountability, transparency, and legitimacy of GenAI systems demands investment in mechanisms that can assess the degree of novelty and trace the statistical lineage of generated outputs. This responsibility must be shared by all stakeholders but especially by those who develop and deploy these technologies at scale.

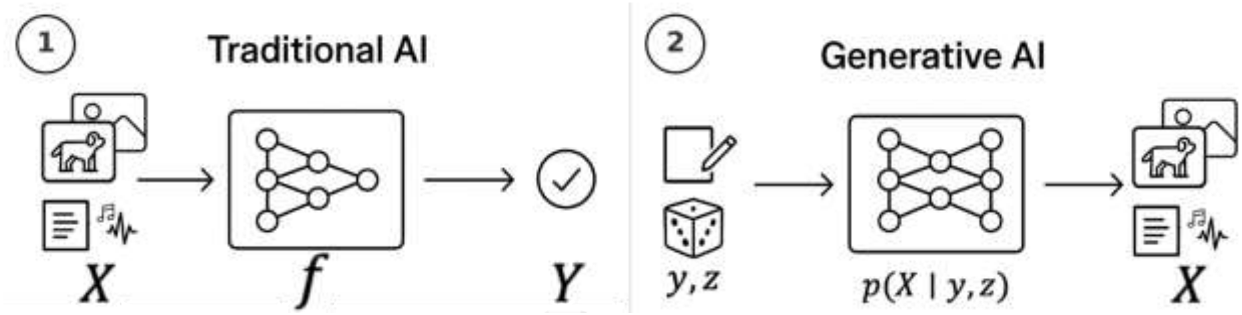
# 1. UNDERSTANDING GENAI SYSTEMS: FROM PREDICTION TO SAMPLING

KEY FINDINGS

- Traditional AI models map input images, text and audio to output labels (e.g., classification), posing minimal copyright risk.
- GenAI models invert this logic: they generate images, text, audio from a distribution, conditioned on abstract prompts and stochastic latent variables.
- Training defines a high-dimensional hypersurface over the output space, shaped by millions of examples.
- The model generates new content by sampling from this hypersurface—implicitly embedding statistical influence from training data.

Generative Artificial Intelligence (GenAI) systems differ fundamentally from traditional machine learning models, as shown in Figure 1. Rather than predicting a label or property from a given input, such as identifying the breed of a dog from an image, GenAI models are trained to synthesize entirely new data points, producing realistic outputs in domains such as language, music, and visual arts. Understanding how these systems work requires shifting from the familiar paradigm of deterministic prediction to one of probabilistic sampling over learned distributions.

Figure 1: (1) Traditional vs. (2) Generative AI: From Prediction to Content Generation



## 1.1. From Human Learning to Statistical Approximation

Unlike human learning, which involves semantic understanding, symbolic reasoning, and contextual generalisation, GenAI operates through statistical approximation. It does not “understand” the content it processes. Instead, it learns regularities, i.e. statistical patterns that appear frequently in large datasets. This distinction is foundational. A GenAI model does not hold the same internal concepts like the humans, but encodes high-dimensional relationships between examples, allowing it to reproduce or extend those patterns probabilistically.

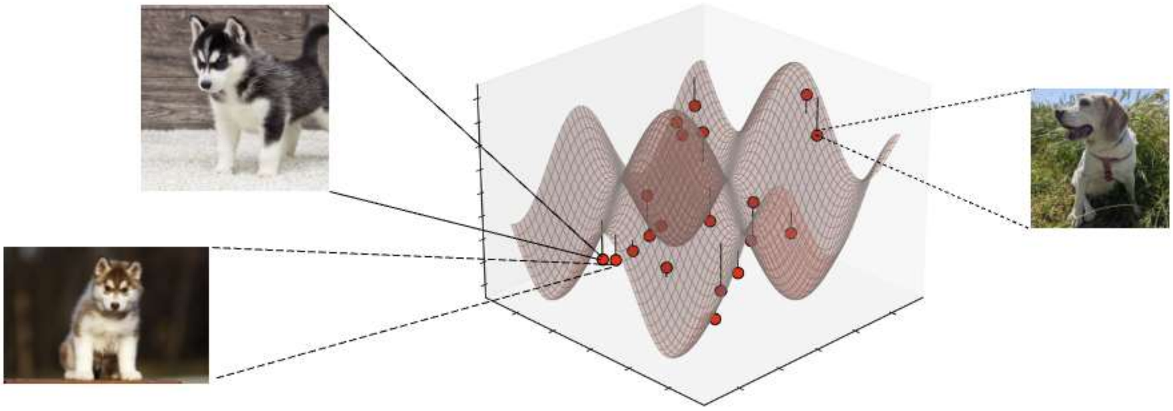
## 1.2. Traditional Predictive Models: The Deterministic Case

In classical supervised learning, a model is trained to learn a mapping function $f: X \rightarrow Y$, where $X$ is the space of inputs and $Y$ is the space of desired outputs. For simplification reasons, we will consider that here this function is typically deterministic[1]: given an input $x$, the model predicts an output $y$. For instance, predicting the sentiment of a review or the species of an animal.

Mathematically, the model constructs a hypersurface in a high-dimensional space, i.e. an approximated mesh that passes near or exactly over the data points in the training set, similar to the one in Figure 2. If the model is too rigid and fits the data exactly, it risks memorisation (i.e., overfitting), failing to generalize beyond the training set. If properly regularized, the model generalizes by smoothing the hypersurface to capture the underlying distribution rather than each example verbatim[2].

The Mesh Metaphor: To aid intuition, imagine the *hypersurface* as a rubber sheet stretched in many dimensions, with each training point exerting a small pull on it. The result is a smooth but complex surface that passes near many of the original data points (the red points in Figure 2).

Figure 2: Visualisation of the AI Model's Learned Hypersurface



This traditional predictive scenario typically does not pose copyright risks because the output variable (e.g. a label or feature) is often *not itself* a protected work (e.g. the dog's age)[3].

---

[1] For clarity and to maintain focus on the core distinctions relevant to this study, classical supervised learning is here presented as a deterministic mapping. In practice, predictive models may embed stochastic components to capture aleatoric uncertainty explicitly (see, e.g., Bishop, 2006). Conversely, in generative modelling, stochasticity is essential: sampling variability is a mathematical requirement to approximate distributions over plausible outputs rather than predicting point estimates. A detailed treatment of probabilistic frameworks exceeds the scope of this analysis.

[2] See Feldman, V. (2020, June). Does learning require memorization? A short tale about a long tail. In Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (pp. 954-959).

[3] In a traditional predictive system, one or more dimensions of this hypersurface (shown here as a 3D mesh) would correspond to output features. For example, in Figure 2, the vertical axis could represent the predicted attribute, such as the estimated age of the dog.

## 1.3. Inversion and the Need for Generative Modelling

Now consider the inverse problem: instead of predicting a property from an input (as show in left (1) part of Figure 1), we want to generate a plausible output (e.g. image, sentence) that corresponds to a general concept (as show in right (2) part of Figure 1).

For instance, rather than classifying the race of a dog based on its picture, we want the model to generate **dogs'** images. The space of valid outputs is vast. There is no single "correct" image of a dog. Hence, we need to model not a deterministic function but a distribution over plausible outputs, i.e., a probabilistic function $p(X \mid y, z)$, where $z$ is a latent random variable that introduces stochastic variability, allowing to approximate all the potential dogs simultaneously. This shift requires generative models to learn conditional probability distributions, which allow for multiple valid outputs to be sampled from the same learned hypersurface.

In the generative setting, the model defines a high-dimensional hypersurface over the output space $X$, which is conditioned on both an external input $y$ (e.g. the race of the dog image to be generated) and a latent random variable $z$. During training, this surface is iteratively deformed to approximate the probability distribution $p(X \mid y, z)$, using millions or billions of training data points. Conceptually, the hypersurface can be viewed as a flexible manifold that bends and stretches in high-dimensional space, adjusting its shape in response to each training example. Each data point exerts a small, localised influence that contributes to the overall geometry of the surface, allowing the model to generalize beyond the observed data while retaining the statistical structure of the training distribution[4].

Once training is complete, generation involves sampling a point on this mesh. This is done by specifying a location in the latent space (typically conditioned on a prompt or noise input) and evaluating the learned function at that point to obtain a data sample.

---

[4] See Chang, Z., Koulieris, G. A., Chang, H. J., & Shum, H. P. (2025). On the design fundamentals of diffusion models: A survey. *Pattern Recognition*, 111934.

# 2. NOVELTY IN HIGH-DIMENSIONAL GENERATIVE SPACES: BETWEEN STATISTICAL EXPLORATION AND STOCHASTIC PARROTING

KEY FINDINGS

- The statistical behaviour of GenAI in high-dimensional spaces undermines conventional metrics of similarity, due to phenomena such as the curse of dimensionality.
- Apparent novelty in outputs may conceal statistical reproduction, particularly in densely represented regions of the generative space.
- The phenomenon of "stochastic parroting" describes instances in which the model inadvertently reproduces training data with high fidelity under the guise of originality.
- Existing similarity assessment tools (e.g. cosine distance, perceptual hashing) lack robustness in determining substantial similarity or derivation.
- Novelty in the context of generative systems must be reinterpreted as a probabilistic property, with implications for copyright qualification and enforcement.

Generative models operate in extremely high-dimensional spaces[5]. In such settings, each data point is represented as a vector in a space that may range from hundreds to tens of thousands of dimensions. For instance, in language models, each token[6] is typically embedded in a vector of 768 to 16,384 dimensions, depending on the architecture. The generative hypersurface learned during training spans this space and encodes statistical regularities extracted from the data. The sheer dimensionality of this space introduces several counterintuitive phenomena that directly affect our ability to assess whether an output is truly novel or, on the contrary, if it is simply copying some training data point[7].

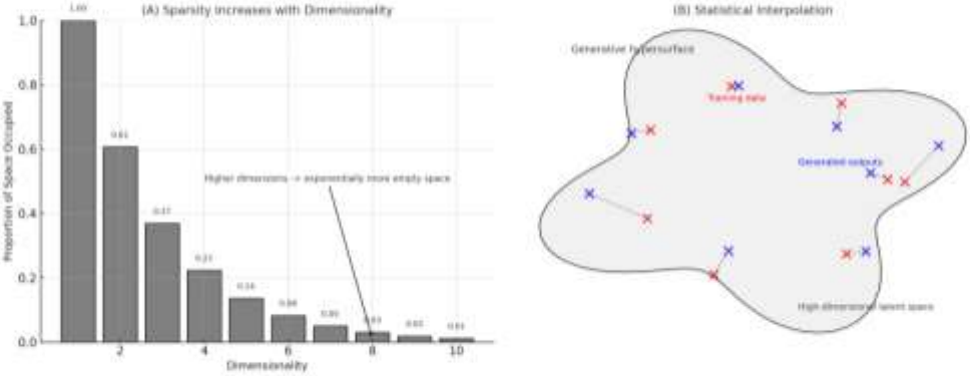
---

[5] In mathematical terms, a "dimension" refers to an axis along which data can vary independently. A "vector" is simply an ordered list of numerical values, each locating the point along these dimensions. While we can only visualise up to three dimensions, vector spaces in AI often have hundreds or thousands, each corresponding to a learned feature. This high dimensionality enables models to capture rich patterns needed to generate realistic text, images, or audio.

[6] A token is a piece of text, such as a word or part of a word, used as the basic unit in language processing. For example, "playing" can be split into the tokens "play" and "ing".

[7] See Gaffar, H., & Albarashdi, S. (2025). Copyright protection for AI-generated works: Exploring originality and ownership in a digital landscape. Asian Journal of International Law, 15(1), 23-46.

Figure 3: Assessing Proximity in High-dimensional GenAI Models



From a mathematical standpoint, high-dimensional spaces exhibit a phenomenon known as the "*curse of dimensionality*", represented in Figure 3. As the number of dimensions increases, the volume of the space grows exponentially. As a consequence, most points in the space become equidistant from one another, and traditional geometric intuitions, such as local proximity or density, lose meaning. A small perturbation in one dimension may result in a large change in the overall vector distance, or conversely, a large perceptual difference may correspond to a minimal displacement in the embedding space. This severely complicates any effort to assess how "close" a generated output is to its nearest training data points.

In practical terms, this means that generative models require an extraordinarily large amount of data, e.g. trillions of points, to adequately approximate the true underlying distribution. Even then, there will be regions of the hypersurface where data is sparse, and others where it is densely clustered. In these dense regions, the surface may be disproportionately influenced by specific training examples[8]. When the model is later queried in these regions, it may generate outputs that strongly reflect those examples, even if they are not exact replicas.

This behaviour is often referred to as *"stochastic[9] parroting"*: It describes the phenomenon where the model, statistically reproduces fragments, patterns, or entire structures that are highly similar or identical to those found in the training set.

Here one of the core challenge lies in distinguishing genuine novelty from these stochastic echoes. Current methods for evaluating originality often rely on similarity metrics such as cosine distance in embedding space[10] or perceptual hashes for images[11]. However, these approaches struggle to provide

---

[8] See Antoniades, A., Wang, X., Elazar, Y., Amayuelas, A., Albalak, A., Zhang, K., & Wang, W. Y. Generalization vs. Memorization: Tracing Language Models' Capabilities Back to Pretraining Data. In *ICML 2024 Workshop on Foundation Models in the Wild*.

[9] Stochastic means probabilistic; it refers to processes involving random variation rather than deterministic rules.

[10] Cosine distance in embedding space measures how similar two vectors are by calculating the angle between them, regardless of their length. In GenAI models, each piece of content is mapped to a vector in a high-dimensional space, and cosine distance helps estimate how close two pieces of content are in terms of learned features.

[11] Perceptual hashes for images create a compact fingerprint that captures visual characteristics, so that similar-looking images have similar hash values even if their raw data differs. This allows approximate matching of content by comparing hash similarity.

robust thresholds for originality, especially because generative models build a complex hypersurface that blends patterns from all the training data, combined with the semantic complexity of creative works. Moreover, the lack of traceability and transparency in model internals further limits our ability to audit or verify the novelty of a given output.

This has direct implications for copyright analysis. A generated work may appear to be new, yet it may lie so close, statistically and semantically, to a protected work that it raises legitimate concerns of derivation. Conversely, a work that seems familiar may not correspond to any specific training example, but rather be a plausible interpolation shaped by multiple influences.

The phenomenon of apparent novelty in generative models must be interpreted within the context of high-dimensional interpolation. While outputs may differ at the surface level —that is, in their visible form or wording— they are statistically entangled with the training data used to construct the generative hypersurface. This means that the distinction between a genuinely new creation and a statistically reproduced artefact is not binary but probabilistic.

Critically, this uncertainty has direct implications for the attribution of authorship and the recognition of creative influence. If the output of a model cannot be functionally disentangled from the data that shaped it, then the line between derivation and originality becomes blurred not only in legal terms but in statistical terms as well.

This raises the next key technical challenge: even if we accept that training data has probabilistically influenced an output, can we identify which data points were most responsible for that influence? In other words, is it possible to trace the statistical lineage of a generated work back to specific works in the training corpus, in a manner that could support reliable attribution?

The following section addresses this challenge in depth, examining the technical limitations of current systems with respect to attribution, the concept of "traceability gaps," and the state of emerging research into probabilistic and functional attribution methods.

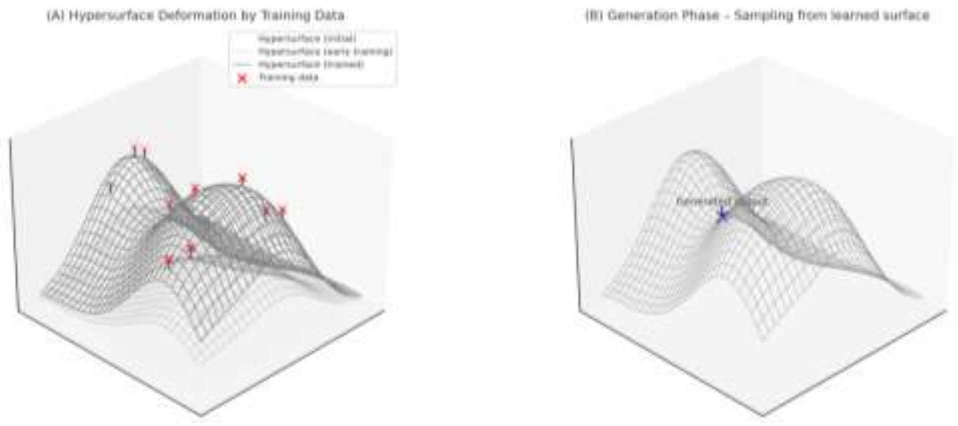
# 3. ATTRIBUTION, TRACEABILITY, AND THE RECOGNITION OF INFLUENCE IN GENAI

KEY FINDINGS

- GenAI architectures do not embed traceability by design, resulting in a structural disconnect between training inputs and generated outputs.
- Attribution of influence remains technically possible in theory but is currently impeded by scalability challenges and the absence of integrated tracing mechanisms.
- Techniques such as influence functions or membership inference remain at a proof-of-concept stage and are not mature for application to large-scale models.
- The lack of attribution undermines legal enforceability, particularly with respect to derivative works and fair remuneration schemes.
- Attribution should evolve from a binary notion to a statistical measure of influence, necessitating new models of accountability and documentation.

One of the most fundamental limitations of current GenAI architectures lies in their inability to provide transparent information about the influence of specific training data on individual outputs. While the underlying hypersurface structure is mathematically shaped by all data points observed during training, there is no internal mechanism that records or preserves a traceable lineage between a generated output and the subset of data that most influenced its creation.

Figure 4: Training Data Shaping the Hypersurface and Sampling without Traceability



## 3.1. The Traceability Gap

This limitation creates what may be termed a traceability gap: a structural disconnect between training inputs and generated outputs. This gap has three main technical causes:

- Stochastic training regimes: Training involves randomised sampling of mini-batches (small subsets of data selected randomly in each step) and stochastic gradient descent (an iterative method that updates model parameters[12] using those subsets), making the learning trajectory path-dependent and not easily reversible.

- Parameter entanglement: In large-scale models, billions of parameters are jointly optimized. Each parameter reflects the aggregated influence of many data points, and no single output can be easily traced to a subset of these.
- Lack of attribution layers: Unlike bibliographic citations in scholarly writing, GenAI systems lack any mechanism, structural or probabilistic, for assigning credit or influence scores to specific training inputs.

Once training concludes, the original data is embedded, and the model is deployed as a functional object, as shown in part (B) of Figure 4. The system retains the parameterised function it learned. Thus, when an output is generated, it is not easy to determine whether it reflects the influence of any specific work, even if that work played a critical role during training.

## 3.2. Probabilistic Influence is Not Observable

From a statistical and functional standpoint, each training example deforms the hypersurface locally and incrementally. The aggregated effect of these deformations defines the generative function. However, due to the non-linear nature of the optimisation process, the relative influence of any one data point is diffuse and distributed. Technically, this is known as the problem of influence attribution[13,14].

Recent efforts in machine learning research have begun to explore methods such as:

- Influence functions: Tools that approximate the marginal effect of a data point on the learned parameters. Generically, these require strong assumptions (e.g., convexity) and are not scalable to modern large-scale transformers[15].
- Membership inference: Statistical tests that assess whether a particular data point was included in the training set. While useful for privacy audits, these methods cannot measure *degree of influence*.
- Attribution audits via synthetic counterfactuals, unlearning or similarity-based approaches: Attempts to remove classes of data post hoc and assess performance degradation. However, in general, such techniques are computationally expensive[16].

---

[12] "Parameters" are the internal numeric values a model learns during training that determine how it produces outputs.

[13] Mlodozeniec, B. K., Eschenhagen, R., Bae, J., Immer, A., Krueger, D., & Turner, R. E. Influence Functions for Scalable Data Attribution in Diffusion Models. In The Thirteenth International Conference on Learning Representations.

[14] Wang, Z., Chen, C., Zeng, Y., Lyu, L., & Ma, S. (2023). Where did I come from? origin attribution of ai-generated images. Advances in neural information processing systems, 36, 74478-74500.

[15] See Koh, P. W., & Liang, P. (2017, July). Understanding black-box predictions via influence functions. In *International conference on machine learning* (pp. 1885-1894). PMLR.

[16] See Choi, W., Koo, J., Cheuk, K. W., Serrà, J., Martínez-Ramírez, M. A., Ikemiya, Y., Murata, N., Wei-Hsiang, L., & Mitsufuji, Y. (2025). Large-Scale Training Data Attribution for Music Generative Models via Unlearning. arXiv preprint arXiv:2506.18312.

None of these methods currently easily scale to models with hundreds of billions of parameters and training sets composed of hundreds of terabytes of heterogeneous data[17]. More importantly, they do not yield definitive answers about the role of specific copyrighted works in shaping a particular output as we will state in the following subsection.

## 3.3. Implications for Copyright Contexts

In the context of copyright law, the inability to trace influence raises serious concerns[18]:

- It undermines the ability of rightsholders to assert whether their works were used to produce a given output.
- It complicates any effort to establish derivative use or non-authorised adaptation under existing legal definitions.
- It makes any mechanism of fair remuneration or licensing dependent on aggregate estimates rather than direct proof.

Moreover, when considering the Text and Data Mining (TDM) exception introduced in the Digital Single Market Directive (Directive (EU) 2019/790)[19], traceability becomes even more relevant. If models trained under this exception later generate outputs that substantially resemble copyrighted works, but attribution cannot be established, legal enforcement becomes practically infeasible.

## 3.4. Current Initiatives and Outlook

While current generative models lack native capabilities for attribution, early-stage initiatives are exploring ways to overcome this limitation[20]. One such example is the case of GEMA, the German society for musical authors and rightsholders. Although their technological approach is still in the pilot phase and not publicly disclosed, their efforts aim precisely to address the issue of attribution at scale, offering mechanisms that may allow rightsholders to assess whether their work has been used during training[21,22].

---

[17] See Park, S. M., Georgiev, K., Ilyas, A., Leclerc, G., & Madry, A. (2023, January). TRAK: Attributing Model Behavior at Scale. In *ICML*.

[18] We carefully recommend revising Abbott, R., & Rothman, E. (2023). Disrupting creativity: Copyright law in the age of generative artificial intelligence. *Fla. L. Rev.*, *75*, 1141.

[19] See Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC, Official Journal of the European Communities 2019 L 130, 92

[20] See Stackpole, B. (2025, March 3). *Bringing transparency to the data used to train artificial intelligence*. MIT Sloan Management Review. https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence

[21] See the GEMA. (2024, November 13). Suno AI and OpenAI: GEMA sues for fair compensation. Retrieved from https://www.gema.de/en/news/ai-and-music/ai-lawsuit

[22] See CISAC. (2025, January 21). *Fair remuneration demanded: GEMA files lawsuit against Suno Inc.* Retrieved from https://www.cisac.org/Newsroom/society-news/fair-remuneration-demanded-gema-files-lawsuit-against-suno-inc

GEMA's work illustrates an essential point: the difficulty of attribution should not be treated as a fixed or intrinsic limitation of GenAI. Rather, it is a contingent technical challenge, solvable in principle through sustained research and institutional support. The lack of attribution mechanisms today should not be interpreted as a justification for non-remunerated use of creative works, but rather as a call to action to invest in the development of auditable and functionally explainable generative systems.

Complementing these early initiatives, recent research[23,24] has systematically documented how the lack of standardized data provenance practices has contributed to the opacity of foundation model training pipelines, highlighting that common tools for tracing data authenticity, consent, and licensing are insufficient to support responsible model development at scale. Their large-scale analysis underscores that without infrastructure for documenting and verifying dataset provenance, efforts to establish attribution and accountability in GenAI systems will remain severely constrained.

These perspectives motivate the following section, which outlines concrete technical recommendations to advance attribution-aware and transparency-enabling infrastructures in GenAI, supporting a more balanced and legally accountable ecosystem for creative rights in the age of artificial content.

---

[23] See Longpre, S., Mahari, R., Obeng-Marnu, N., Brannon, W., South, T., Gero, K. I., & Kabbara, J. (2024, July). Position: Data Authenticity, Consent, & Provenance for AI are all broken: what will it take to fix them?. In Forty-first International Conference on Machine Learning.

[24] See Rights Alliance. (2024). Report on AI model providers' training data transparency and enforcement of copyrights. https://rettighedsalliancen.dk/wp-content/uploads/2024/09/Report-on-AI-model-providers-training-data-transparency-and-enforcement-of-copyrights.pdf

# 4. TECHNICAL RECOMMENDATIONS TO ENHANCE TRANSPARENCY AND LEGAL COMPLIANCE IN GENAI

KEY FINDINGS

- Technical solutions to traceability and attribution challenges are feasible but require targeted research, standardisation, and industry adoption and investment.
- Developers of GenAI systems should bear proactive responsibility for documenting dataset provenance and enabling model-level transparency audits.
- Remuneration frameworks should reflect not only literal copying but also statistical usage and cumulative influence over generated outputs.
- Open standards, auditable infrastructures, and structured collaboration between rightsholders, researchers, and regulators are critical to ensure compliance and legitimacy.
- The European Union is well positioned to lead the establishment of a balanced regulatory-technical ecosystem that safeguards creators while enabling responsible innovation.

The preceding sections have outlined the technical foundations that make attribution and novelty assessment particularly difficult in the context of generative models. These challenges are not intrinsic limitations of artificial intelligence, but rather the result of current design choices and insufficient investment in transparency-focused infrastructure. Addressing these gaps requires deliberate technical commitments from developers and broader policy support. This section proposes a set of evidence-based recommendations focused on enabling better governance, accountability, and legal compliance in GenAI systems.

## 4.1. Responsibility Must Be Assumed by All the Stakeholders

The burden of addressing attribution and novelty should not rest solely on rightsholders or public institutions. Developers of GenAI systems, particularly those operating at industrial scale, must be expected to invest in mechanisms that make the models auditable, traceable, and assessable in terms of their statistical dependencies on copyrighted data. Just as the use of personal data in algorithmic systems now requires formal risk assessments and technical safeguards, the use of creative works must be governed by a comparable framework of technical due diligence.

This includes:

- Implementing systems to assess whether training datasets contain copyrighted content and documenting such assessments.
- Developing internal capabilities for attribution inference, influence scoring, or membership estimation.
- Providing technical documentation and disclosures that enable third-party auditing and legal verification.

## 4.2. From Literal Copying to Statistical Use: Rethinking Remuneration Models

Current frameworks for copyright enforcement often focus on detecting literal copying or close derivations. However, as discussed above, generative systems function by learning statistical approximations of input data, not storing it directly. As such, the notion of "use" in GenAI must be broadened to include functional dependence and distributional influence.

A combination of technical, legal, and economic approaches is needed to account for:

- The fact that models can integrate statistical characteristics of protected works, even when no output is a verbatim copy.
- The difficulty of identifying which portions of large-scale datasets materially affect specific regions of the generative space.
- The cumulative value created by aggregating countless small inputs (e.g., millions of short text snippets or musical phrases) that individually appear insignificant.

In this context, remuneration schemes should not be limited to literal matches, but should explore mechanisms based on *statistical usage*, such as:

- Pool-based royalties indexed on dataset inclusion.
- Contribution-weighted compensation models, derived from influence estimation tools.
- Public registries of data sources and opt-in compensation systems for creators.

## 4.3. Traceability Infrastructure: A Research and Governance Priority

Technical progress in attribution and traceability is not only possible, but already underway. Emerging methods such as membership inference, dataset influence functions, and semantic similarity detection offer a growing foundation for tracing the statistical impact of individual data points on model outputs.

Public investment and standardisation efforts should prioritise:

- Open-source frameworks for auditing GenAI training pipelines[25].
- Tools for measuring the likelihood that a specific data point influenced a generated sample.
- Standard protocols for dataset documentation and provenance tagging.
- Development of independent test suites for evaluating traceability and novelty across models.

---

[25] See Zhong, H., Chang, J., Yang, Z., Wu, T., Mahawaga Arachchige, P. C., Pathmabandu, C., & Xue, M. (2023, April). Copyright protection and accountability of generative ai: Attack, watermarking and attribution. In *Companion Proceedings of the ACM Web Conference 2023* (pp. 94-98).

## 4.4. Standardisation, Auditing, and Cooperation with Rightsholders

Establishing technical standards is critical to ensure interoperability and accountability in the GenAI ecosystem. Standards should:

- Define what constitutes sufficient documentation of training sources.
- Provide specifications for model-level attribution reporting.
- Support audit mechanisms that are usable by rightsholders and institutions.

Beyond standards and regulations[26,27], **cooperation between GenAI developers, authors' societies,** academic researchers, and regulators must be actively encouraged. Collaborative frameworks can:

- Facilitate the safe sharing of annotated datasets for testing attribution methods.
- Align research incentives with regulatory needs.
- Ensure that creators have access to meaningful and usable tools to protect their rights.

This set of recommendations provides a technically grounded pathway for improving transparency, attribution, and fairness in generative AI[28]. Without such measures, the growing asymmetry between powerful GenAI models and the rights of original creators risks undermining the legitimacy of the digital content economy.

---

[26] See European Commission. (2024). *Artificial Intelligence Act – Recital 107*. https://ai-act-law.eu/recital/107/

[27] See Banterle, F., & Schettino, A. (2024). Copyright provisions in the AI Act: Generative AI, transparency, and data mining. https://www.hoganlovells.com/en/publications/copyright-provisions-in-the-ai-act-generative-ai-transparency-and-data-mining

[28] Aligned with the OECD (2024), "AI, data governance and privacy: Synergies and areas of international co-operation", *OECD Artificial Intelligence Papers*, No. 22, OECD Publishing, Paris, https://doi.org/10.1787/2476b1a4-en.

# 5. CONCLUSION

The current capabilities of generative artificial intelligence (GenAI) models are grounded in statistical approximation and high-dimensional function learning. These systems do not operate through inexplicable means but rely on probabilistic interpolations derived from vast training datasets. While the internal representations they build, often in the form of complex hypersurfaces, are not directly interpretable, they are ultimately governed by mathematical principles, and thus susceptible to systematic evaluation, refinement, and regulation.

The technical limitations we have described (see Section 1-3), particularly those related to memorisation, novelty, and traceability, are not beyond resolution, but they require dedicated investment, methodological innovation, and proactive governance. Ignoring these issues risks undermining the intellectual property ecosystem and public trust in the responsible development of AI systems.

Traditional notions of authorship, influence, and originality require reinterpretation in the light of stochastic generation and high-dimensional modelling. This calls for a cross-sectoral response that integrates expertise from AI research, copyright law, and ethical governance (see Section 4).

GenAI systems open transformative possibilities for creativity and productivity, but they also pose new and subtle risks. Their power lies not in being autonomous or "intelligent" in the human sense, but in their capacity to amplify and remix statistical structures learned from existing content. This makes them both extraordinarily valuable and inherently entangled with prior human work.

The European Union is uniquely positioned to lead the way in defining normative and technical benchmarks for transparency, attribution, and responsibility in GenAI. Doing so will not only protect rightsholders and democratic values, but will also foster innovation grounded in legitimacy, accountability, and social trust.

# REFERENCES

- Abbott, R., & Rothman, E. (2023). Disrupting creativity: Copyright law in the age of generative artificial intelligence. Fla. L. Rev., 75, 1141
- Antoniades, A., Wang, X., Elazar, Y., Amayuelas, A., Albalak, A., Zhang, K., & Wang, W. Y. Generalization vs. Memorization: Tracing Language Models' Capabilities Back to Pretraining Data. In ICML 2024 Workshop on Foundation Models in the Wild.
- Banterle, F., & Schettino, A. (2024). Copyright provisions in the AI Act: Generative AI, transparency, and data mining. https://www.hoganlovells.com/en/publications/copyright-provisions-in-the-ai-act-generative-aitransparency-and-data-mining
- Bishop, C. M. (2006). Pattern Recognition and Machine Learning. New York: Springer.
- Chang, Z., Koulieris, G. A., Chang, H. J., & Shum, H. P. (2025). On the design fundamentals of diffusion models: A survey. Pattern Recognition, 111934.
- Choi, W., Koo, J., Cheuk, K. W., Serrà, J., Martínez-Ramírez, M. A., Ikemiya, Y., Murata, N., Wei-Hsiang, L., & Mitsufuji, Y. (2025). Large-Scale Training Data Attribution for Music Generative Models via Unlearning. arXiv preprint arXiv:2506.18312.
- CISAC. (2025, January 21). Fair remuneration demanded: GEMA files lawsuit against Suno Inc. Retrieved from https://www.cisac.org/Newsroom/society-news/fair-remuneration-demanded-gema-files-lawsuit-against-suno-inc
- Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on Copyright and Related Rights in the Digital Single Market and Amending Directives 96/9/EC and 2001/29/EC, Official Journal of the European Communities 2019 L 130, 92
- European Commission. (2024). Artificial Intelligence Act – Recital 107. https://ai-act-law.eu/recital/107/
- Feldman, V. (2020, June). Does learning require memorization? a short tale about a long tail. In Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing (pp. 954-959).
- Gaffar, H., & Albarashdi, S. (2025). Copyright protection for AI-generated works: Exploring originality and ownership in a digital landscape. Asian Journal of International Law, 15(1), 23-46.
- GEMA. (2024, November 13). Suno AI and OpenAI: GEMA sues for fair compensation. Retrieved from https://www.gema.de/en/news/ai-and-music/ai-lawsuit
- Koh, P. W., & Liang, P. (2017, July). Understanding black-box predictions via influence functions. In International conference on machine learning (pp. 1885-1894). PMLR.
- Longpre, S., Mahari, R., Obeng-Marnu, N., Brannon, W., South, T., Gero, K. I., ... & Kabbara, J. (2024, July). Position: Data Authenticity, Consent, & Provenance for AI are all broken: what will it take to fix them?. In Forty-first International Conference on Machine Learning.

- Mlodozeniec, B. K., Eschenhagen, R., Bae, J., Immer, A., Krueger, D., & Turner, R. E. Influence Functions for Scalable Data Attribution in Diffusion Models. In The Thirteenth International Conference on Learning Representations.
- OECD (2024), "AI, data governance and privacy: Synergies and areas of international cooperation", OECD Artificial Intelligence Papers, No. 22, OECD Publishing, Paris, https://doi.org/10.1787/2476b1a4-en.
- Park, S. M., Georgiev, K., Ilyas, A., Leclerc, G., & Madry, A. (2023, January). TRAK: Attributing Model Behavior at Scale. In ICML.
- Rights Alliance. (2024). Report on AI model providers' training data transparency and enforcement of copyrights. https://rettighedsalliancen.dk/wp-content/uploads/2024/09/Report-on-AI-model-providers-training-datatransparency-and-enforcement-of-copyrights.pdf
- Stackpole, B. (2025, March 3). Bringing transparency to the data used to train artificial intelligence. MIT Sloan Management Review. https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-trainartificial-intelligence
- Wang, Z., Chen, C., Zeng, Y., Lyu, L., & Ma, S. (2023). Where did I come from? origin attribution of ai-generated images. Advances in neural information processing systems, 36, 74478-74500.
- Zhong, H., Chang, J., Yang, Z., Wu, T., Mahawaga Arachchige, P. C., Pathmabandu, C., & Xue, M. (2023, April). Copyright protection and accountability of generative ai: Attack, watermarking and attribution. In Companion Proceedings of the ACM Web Conference 2023 (pp. 94-98).

This in-depth analysis explains the statistical nature of generative AI and how training on copyright-protected data results in persistent functional dependencies with respect to the used data. It highlights the challenges of attribution and novelty detection in these high-dimensional models, emphasising the limitations of current methodologies. The study provides technical recommendations for traceability and output assessment mechanisms.

This study is commissioned by the European Parliament's Policy Department for Justice, Civil Liberties and Institutional Affairs at the request of the Committee on Legal Affairs.

PE 776.529

Print ISBN 978-92-848-2903-3 | doi:10.2861/4727951 | QA-01-25-156-EN-C
PDF ISBN 978-92-848-2902-6 | doi:10.2861/4255296 | QA-01-25-156-EN-N