BRIEFING

Requested by the JURI Committee

European Parliament

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

简报由JURI委员会请求

European Parliament

生成式人工智能在版权背景下的技术方面

生成式人工智能超曲面中的归属与新颖性

大法官政策部门、公民自由与机构事务司、欧洲议会公民权利、大法官与机构事务总司
PE 776.529 - 2025年7月

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.

生成式人工智能在版权背景下的技术方面

生成式 AI 超曲面中的归因与新颖性

摘要

本深入分析解释了生成式人工智能的统计性质，以及在版权保护数据上训练如何导致与所用数据相关的持续功能依赖。它强调了在这些高维模型中归因和新颖性检测的挑战，突出当前方法的局限性。该研究为可追溯性和输出评估机制提供了技术建议。本研究由大法官、民权与机构事务政策部门应法律事务委员会的请求进行。

commissioned by the European Parliament's

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

This document is available on the internet at:
http://www.europarl.europa.eu/supporting-analyses

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

作者
Dr Axel BRANDO，HPES 实验室，巴塞罗那超级计算中心

ADMINISTRATOR RESPONSIBLE
Mariusz MACIEJEWSKI

EDITORIAL ASSISTANT
Ivona KLECAN

语言版本原文：EN

关于编辑
政策部门提供内部和外部专业知识，以支持欧洲议会委员会及其他议会机构制定立法并对欧盟内部政策进行民主监督。

如需联系大法官政策部或订阅更新，请写信至：大法官、民权与机构事务
政策部 欧洲议会 B-1047 布鲁塞尔 电子邮件：
poldep-iust-b@europarl.europa.eu

本文件可在互联网上获取：
http://www.europarl.europa.eu/supporting-analyses

CONTENTS

目录

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 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

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.

EXECUTIVE SUMMARY

生成式人工智能（GenAI）系统代表了数字内容生产方式的根本转变。与传统的预测系统不同，后者旨在推断输入的属性或特征（通常不涉及版权保护材料），GenAI模型被设计用于在文本、图像和音乐等领域生成类人内容；这些领域通常涉及知识产权。

从技术上讲，GenAI模型可以被理解为一个高维超曲面，是由训练数据点形成空间内的可变形网格。在训练过程中，该网格逐步变形以逼近数据集中的点，理想情况下接近或在某些情况下精确通过这些点，同时避免强烈或不规则的弯曲。这种平滑的变形使模型能够超越训练数据进行泛化。训练完成后，生成内容意味着在特定坐标处评估该超曲面。该采样过程可能产生看似全新的输出，也可能产生与原始训练数据无法区分或完全相同的输出。

核心技术挑战在于训练数据与生成时学习到的超曲面之间存在的函数依赖关系。由于生成输出对应于选择曲面上的一个点，而该曲面本身是由训练数据塑造的，因此任何输出与影响它的原始数据之间的关系并非直接可追溯，但每个生成点都是训练数据累积统计影响导致曲面变形的结果。

probabilistically persists within the model's generative function.

Two key copyright-relevant issues arise from this dependency:

1. 归因：我们如何确定生成的输出是否部分或实质上有意义地源自一个或多个训练示例？当前模型没有可靠的机制来量化或追踪这种影响。

2. 新颖性：我们如何评估一个输出是否构成真正的新创作，还是高维随机鹦鹉学舌的结果，即在新颖性幌子下统计上几乎逐字复制模式？

本研究邀请对GenAI时代影响、派生和原创性的解释进行根本性重新思考。它强调需要重新评估诸如数字单一市场指令（指令（EU）2019/790）中引入的文本和数据挖掘（TDM）例外等法律框架的作用和范围，尤其是在基于此类例外训练的模型生成的输出可能引发版权侵权问题时。

解决这些挑战不仅仅是法律解释的问题；它需要直接的技术参与。确保GenAI系统的问责性、透明度和合法性，要求投资于能够评估新颖程度并追踪生成输出统计来源的机制。这一责任必须由所有利益相关者共同承担，尤其是那些大规模开发和部署这些技术的人。

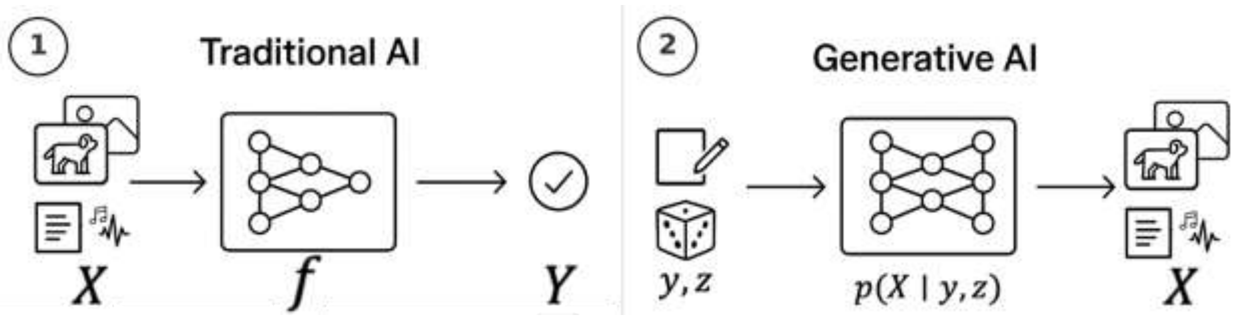
1. UNDERSTANDING GENAI SYSTEMS: FROM PREDICTION TO SAMPLING

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.

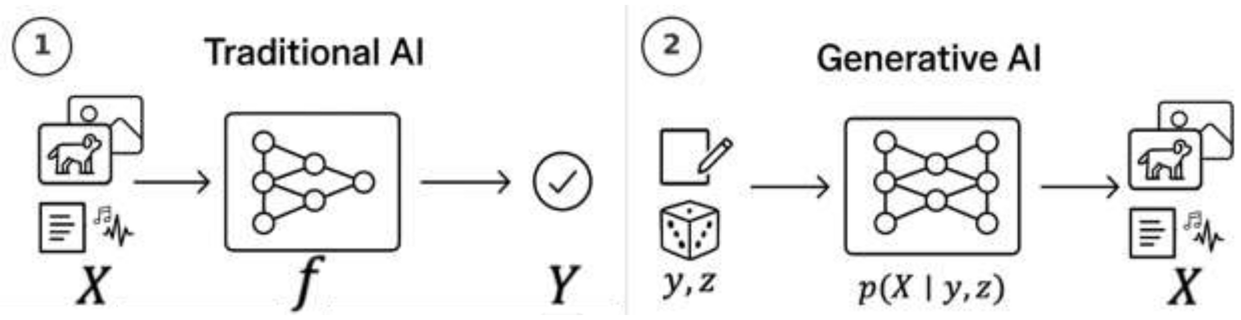
1. UNDERSTANDING GENAI SYSTEMS: FROM PREDICTION TO SAMPLING

KEY FINDINGS

- 传统AI模型将输入的图像、文本和音频映射到输出标签（例如，分类），因此版权风险较小。
- GenAI模型颠倒了这一逻辑：它们根据抽象提示和随机潜变量，从分布中生成图像、文本、音频。
- 训练定义了一个高维超曲面，覆盖输出空间，由数百万个示例塑造。
- 该模型通过从该超曲面采样生成新内容，隐式地嵌入了训练数据的统计影响。

生成式人工智能（GenAI）系统与传统机器学习模型有根本区别，如图1所示。GenAI模型不是从给定输入中预测标签或属性，例如从图像中识别狗的品种，而是被训练来合成全新的数据点，在语言、音乐和视觉艺术等领域生成逼真的输出。理解这些系统的工作原理需要从熟悉的确定性预测范式转变为对学习分布进行概率采样的范式。

图1：（1）传统与（2）生成式人工智能：从预测到内容生成



1.1. 从人类学习到统计近似

与涉及语义理解、符号推理和上下文泛化的人类学习不同，GenAI通过对其处理内容的统计近似来运作。它学习的是规律性，即在大型数据集中频繁出现的统计模式。这一区别是基础性的。It does not “understand” the GenAI模型不具备像人类那样的内部概念，而是编码示例之间的高维关系，使其能够以概率方式再现或扩展这些模式。

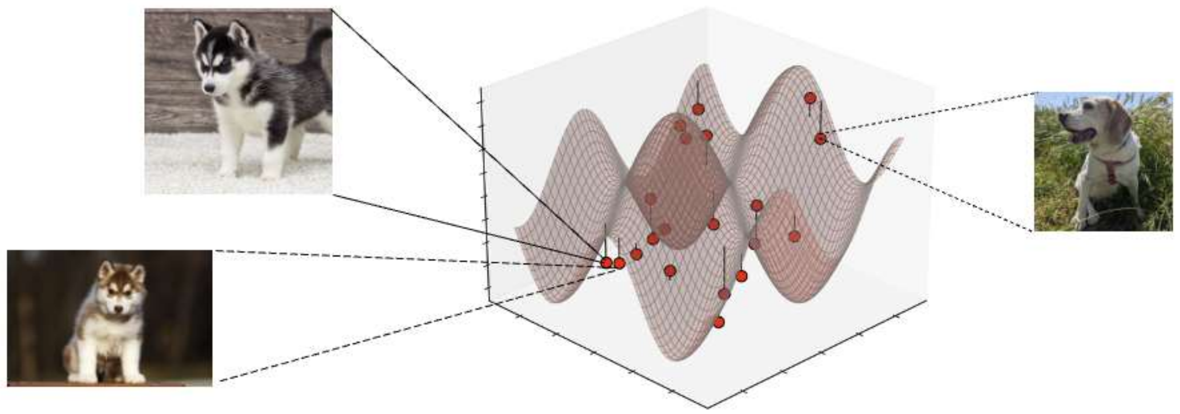
1.2. Traditional Predictive Models: The Deterministic Case

In classical supervised learning, a model is trained to learn a mapping function $f: X \to Y$, 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.

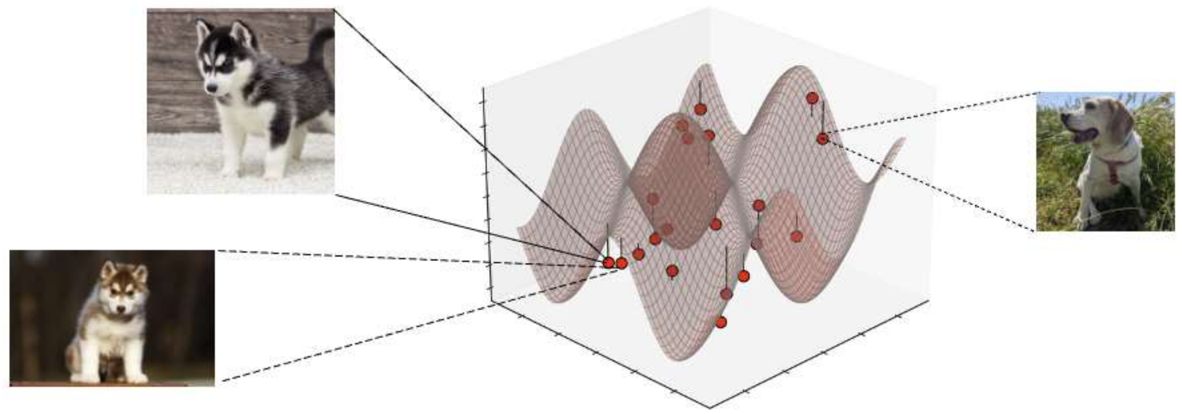
1.2. 传统预测模型：确定性情况

在经典的监督学习中，模型被训练以学习映射函数 $f: X \to Y$，其中 X 是输入空间，Y 是期望输出空间。为简化起见，我们这里将认为该函数通常是确定性的[1]：给定输入 x，模型预测输出 y。例如，预测评论的情感或动物的物种。

从数学上讲，该模型在高维空间中构建了一个超曲面，即一个近似网格，经过训练集中的数据点附近或正好经过这些点，类似于图2中的情况。如果模型过于刚性并且完全拟合数据，则存在记忆化（即过拟合）的风险，无法在训练集之外进行泛化。如果适当正则化，模型通过平滑超曲面来泛化，以捕捉潜在分布，而不是逐字逐句地记忆每个样本[2]。

> 网格隐喻：为了帮助直观理解，可以将超曲面想象成在多维空间中拉伸的橡胶片，每个训练点对其施加一个小的拉力。结果是一个平滑但复杂的曲面，经过许多原始数据点附近（图2中的红点）。

图2：可视化　ation of the AI Model's Learned Hypersurfac



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 为了清晰并保持对本研究相关核心区别的关注，经典监督学习在此被呈现为确定性映射。实际上，预测模型可能嵌入随机成分以显式捕捉偶然性不确定性（参见例如 Bishop, 2006）。相反，在生成建模中，随机性是必需的：采样变异性是数学要求，用于近似可能输出的分布，而非预测点估计。概率框架的详细处理超出了本分析的范围。2 参见 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 在传统预测系统中，该超曲面的一维或多维（此处显示为3D网格）将对应输出特征。例如，在图2中，垂直轴可以表示预测属性，如估计的狗的年龄。

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.

1.3. 反演与生成建模的需求

现在考虑逆问题：不是从输入预测属性（如图1左侧(1)部分所示），而是想生成一个对应于一般概念的合理输出（例如图像、句子）（如图1右侧(2)部分所示）。

例如，我们不再是基于狗的图片来分类其品种，而是希望模型生成图像。因此，我们需要建模的不是确定性函数，而是对合理输出的分布，即一个概率函数 $p(X \mid y, z)$，其中 z 是引入随机变异的潜在随机变量，允许同时逼近所有潜在的狗。这一转变要求生成模型学习条件概率分布，从同一学习到的超曲面中采样多个有效输出。

在生成设置中，模型定义了一个关于输出空间的高维超曲面 X，该超曲面同时以外部输入 y（例如要生成的狗的品种图像）和潜在随机变量 z 为条件。在训练过程中，该曲面通过数百万或数十亿的训练数据点迭代变形，以逼近概率分布 $p(X \mid y, z)$。从概念上讲，该超曲面可视为一个灵活的流形，在高维空间中弯曲和拉伸，根据每个训练样本调整其形状。每个数据点施加一个小的、局部的影响，促成曲面整体几何形状的形成，使模型能够超越观察到的数据进行泛化，同时保留训练分布的统计结构[4]。

训练完成后，生成过程涉及在该网格上采样一个点。这是通过在潜在空间中指定一个位置（通常基于提示或噪声输入进行条件化），并在该点评估学习到的函数以获得数据样本来完成的。

[4] 参见 Chang, Z., Koulieris, G. A., Chang, H. J., & Shum, H. P. (2025). 关于扩散模型设计基础的综述。模式识别, 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.

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

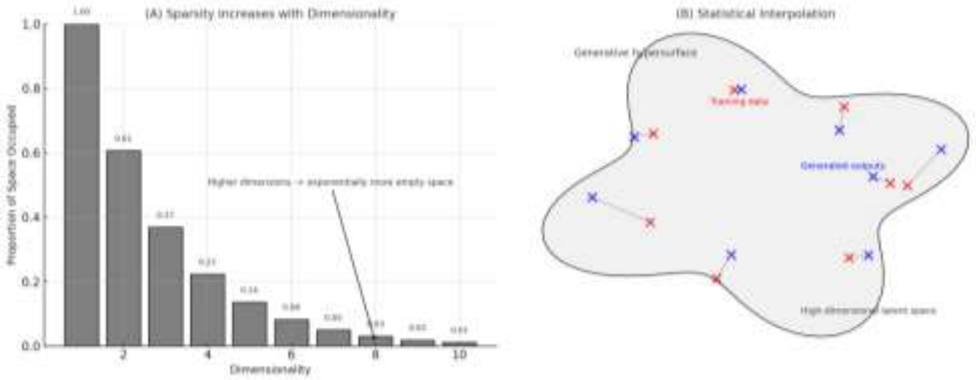
主要发现

- 生成式人工智能在高维空间中的统计行为削弱了传统的相似性度量，这归因于维度灾难等现象。
- 输出中表面的新颖性可能掩盖统计复制，尤其是在生成空间中密集表示的区域。
- 在原创的幌子下无意中高度还原训练数据。The phenomenon of "stochastic parroting" describes instances in which the model
- 现有的相似性评估工具（例如余弦距离、感知哈希）在判断实质性相似性或派生性方面缺乏鲁棒性。
- 在生成系统的背景下，新颖性必须被重新解释为一种概率属性，这对版权资格认定和执行具有影响。

生成模型在极高维空间中运行[5]。在这种环境下，每个数据点被表示为一个向量，空间维度可能从数百到数万不等。例如，在语言模型中，每个标记[6]通常被嵌入到一个768到16,384维的向量中，具体取决于架构。训练期间学习到的生成超曲面跨越该空间，并编码了从数据中提取的统计规律。该空间的高维特性引入了若干反直觉现象，这些现象直接影响我们评估输出是否真正新颖，或者相反，仅仅是复制某些训练数据点[7]的能力。

[5] 有序的数值列表，每个数值定位于这些维度上的一个点。虽然我们只能可视化最多三个维度，但人工智能中的向量空 In mathematical terms, a "dimension" refers to an axis along which data can vary independently. A "vector" is simply an 间通常有数百或数千个维度，每个维度对应一个学习到的特征。这种高维度使模型能够捕捉生成逼真文本、图像或音频所需的丰富模式。[6] 令牌是文本的一部分，如一个词或词的一部分，用作语言处理的基本单位。例如，[7] 参见 Gaffar, H., & Albarashdi, S. (2025). 版权保护人工智能生成作品：探索数字环境中的原创性和所有权。《亚洲国际法杂志》，15(1), 23-46。
"playing" can be split into the tokens "play" and "ing".

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

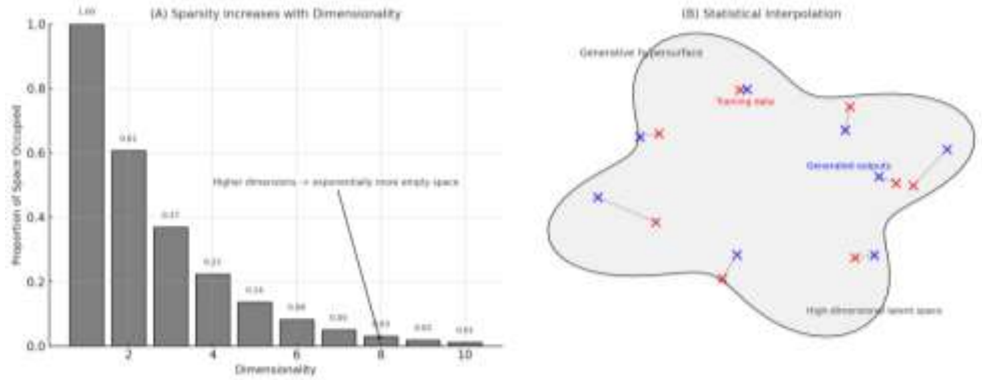
[8] See Antoniades, A., Wang, X., Elazar, Y., Amayuelas, A., Albalak, A., Zhang, K., & Wang, W. Y. Generalization vs. Memorization: Tracing Language Models' Capabilities Back to Pretraining Data. In *ICML 2024 Workshop on Foundation Models in the Wild*.

[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.

图3：评估高维GenAI模型中的接近度



从数学角度来看，高维灾难，如图3所示。随着维度数量的增加，空间的体积呈指数增长。因此，空间中的大多数点彼此距离相等，传统的几何直觉，如局部邻近性或密度，失去意义。一个维度上的微小扰动可能导致整体向量距离的大幅变化，反之，一个大的感知差异可能对应嵌入空间中的最小位移。这极大地复杂化了评估生成输出与其最近训练数据点“接近”程度的任何努力。

在实际操作中，这意味着生成模型需要极其庞大的数据量，例如数万亿个点，才能充分逼近真实的潜在分布。即便如此，超曲面的某些区域数据稀疏，而另一些区域则密集聚集。在这些密集区域，曲面可能会被特定的训练样本不成比例地影响[8]。当模型随后在这些区域被查询时，可能会生成强烈反映这些样本的输出，即使它们不是完全复制品。

随机鹦鹉学舌：描述模型在统计上重现与训练集高度相似或相同的片段、模式或整体结构的现象。

这里的核心挑战之一在于区分真正的新颖性与这些随机回声。当前评估原创性的方法通常依赖于嵌入空间中的余弦距离[10]或图像的感知哈希[11]等相似度度量。然而，这些方法难以提供

[8] 参见 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. 载于ICML 2024 Workshop on Foundation Models in the Wild。[9] 随机意指概率性的；指涉及随机变化而非确定性规则的过程。[10] 嵌入空间中的余弦距离通过计算两个向量之间的夹角来衡量它们的相似度，而不考虑它们的长度。在GenAI模型中，每个内容片段被映射到高维空间中的一个向量，余弦距离有助于估计两个内容在学习特征上的接近程度。[11] 图像的感知哈希创建一个紧凑的指纹，捕捉视觉特征，使得即使原始数据不同，外观相似的图像也具有相似的哈希值。这允许通过比较哈希相似度来近似匹配内容。

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.

稳健的原创性阈值，特别是因为生成模型构建了一个复杂的超曲面，融合了所有训练数据的模式，再加上创意作品的语义复杂性。此外，模型内部缺乏可追溯性和透明度，进一步限制了我们审计或验证给定输出新颖性的能力。

这对版权分析有直接影响。生成的作品可能看似新颖，但在统计和语义上可能与受保护作品非常接近，从而引发合理的衍生性担忧。相反，看似熟悉的作品可能并不对应任何具体的训练示例，而是由多种影响塑造出的合理插值。

生成模型中表面新颖的现象必须在高维插值的背景下进行解读。虽然输出在表面层面——即其可见形式或措辞上可能不同，但它们在统计上与用于构建生成超曲面的训练数据纠缠在一起。这意味着真正的新创作与统计再现的人工制品之间的区别不是二元的，而是概率性的。

关键是，这种不确定性对作者归属和创作影响的认可有直接影响。如果模型的输出无法在功能上与塑造它的数据区分开来，那么衍生性与原创性之间的界限不仅在法律上模糊，在统计上也同样模糊。

这提出了下一个关键的技术挑战：即使我们接受训练数据在概率上影响了输出，我们能否识别出哪些数据点对这种影响负有最大责任？换句话说，是否有可能以一种能够支持可靠归因的方式，将生成作品的统计血统追溯到训练语料库中的具体作品？

以下部分将深入探讨这一挑战，审视当前系统在归因方面的技术限制、“可追溯性缺口”的概念，以及概率和功能归因方法的新兴研究现状。

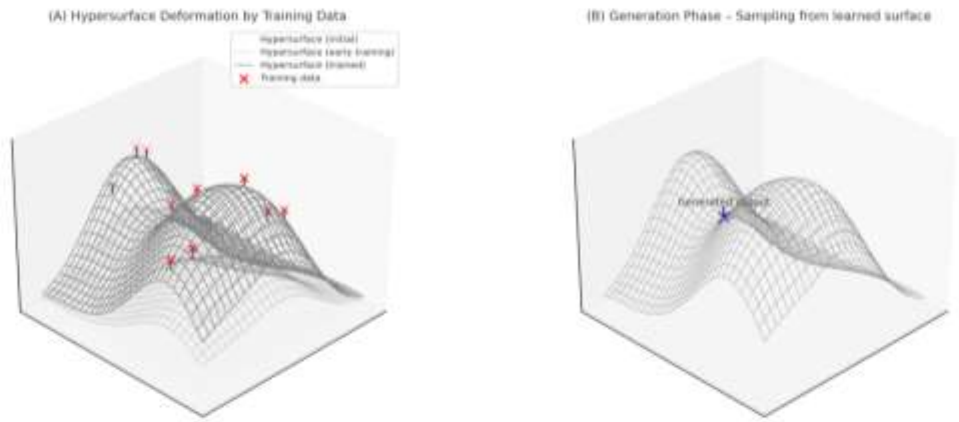
3. ATTRIBUTION, TRACEABILITY, AND THE RECOGNITION OF INFLUENCE IN GENAI

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

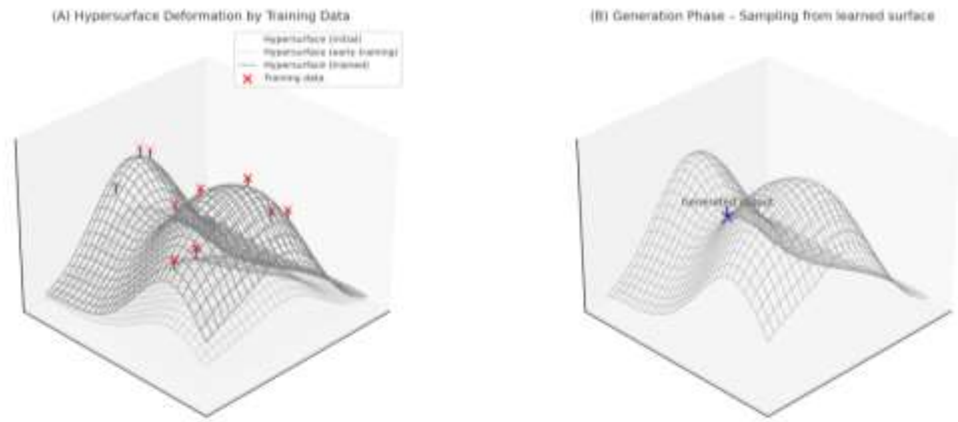
3. ATTRIBUTION, TRACEABILITY, AND THE RECOGNITION OF INFLUENCE IN GENAI

KEY FINDINGS

- GenAI 架构设计上不嵌入可追溯性，导致训练输入与生成输出之间存在结构性断裂。
- 影响归属在理论上技术上仍然可能，但目前受限于可扩展性挑战和缺乏集成追踪机制。
- 诸如影响函数或成员推断等技术仍处于概念验证阶段，尚未成熟到可应用于大规模模型。
- 缺乏归属削弱了法律的可执行性，特别是在衍生作品和公平报酬方案方面。
- 归因应从二元概念演变为影响的统计度量，这需要新的问责和文档模型。

当前GenAI架构最根本的限制之一在于它们无法提供关于特定训练数据对单个输出影响的透明信息。虽然底层的超曲面结构在数学上由训练期间观察到的所有数据点塑造，但没有内部机制记录或保留生成输出与最影响其创建的数据子集之间的可追溯血统。

图4：训练数据塑造超曲面及无可追溯性的采样



3.1. 可追溯性缺口

这一限制造成了所谓的可追溯性缺口：训练输入与生成输出之间的结构性断裂。该缺口有三个主要技术原因：

- 随机训练方案：训练涉及小批量的随机采样（每一步随机选择的小数据子集）和随机梯度下降（迭代

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.

使用那些子集更新模型参数的方法[12]，使得学习轨迹依赖路径且不易逆转。

- 参数纠缠：在大规模模型中，数十亿参数被共同优化。每个参数反映了许多数据点的综合影响，且单个输出无法轻易追溯到这些数据点的子集。
- 缺乏归因层：与学术写作中的书目引用不同，GenAI系统缺乏任何机制，无论是结构性的还是概率性的，用于为特定训练输入分配信用或影响分数。

训练结束后，原始数据被嵌入，模型作为一个功能性对象被部署，如图4的(B)部分所示。系统保留了其学习到的参数化函数。因此，当生成输出时，很难确定其是否反映了任何特定作品的影响，即使该作品在训练过程中起到了关键作用。

3.2. 概率性影响不可观测

从统计和功能的角度来看，每个训练样本都会局部且渐进地变形超曲面。这些变形的累积效应定义了生成函数。然而，由于优化过程的非线性特性，任何一个数据点的相对影响是分散且分布式的。从技术上讲，这被称为影响归因问题[13,14]。

机器学习研究的最新努力已经开始探索以下方法：

- 影响函数：用于近似数据点对学习参数边际影响的工具。一般来说，这些方法需要强假设（例如凸性），且无法扩展到现代大规模transformers[15]。
- 成员推断：统计测试，用于评估特定数据点是否包含在训练集中。虽然对隐私审计有用，但这些方法无法衡量影响程度。
- 通过合成反事实、遗忘或基于相似性的方法进行归因审计：尝试事后移除某类数据并评估性能下降。然而，通常这类技术计算成本较高[16]。

[12] 训练期间决定其如何产生输出。“Parameters” are the internal numeric values a model learns [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

目前这些方法都难以轻松扩展到拥有数千亿参数和由数百TB异构数据组成的训练集的模型[17]。更重要的是，正如我们将在以下小节中说明的，它们无法对特定版权作品在塑造特定输出中的作用给出明确答案。

3.3. 对版权背景的影响

在版权法背景下，无法追踪影响引发了严重的担忧[18]：

- 这削弱了权利人主张其作品是否被用于生成特定输出的能力。
- 这使得根据现有法律定义确定衍生使用或未经授权的改编变得复杂。
- 它使任何公平报酬或许可机制依赖于总体估计而非直接证据。

此外，在考虑数字单一市场指令（指令(EU) 2019/790）[19]中引入的文本和数据挖掘（TDM）例外时，追溯性变得更加重要。如果在该例外下训练的模型后来生成的输出与版权作品实质上相似，但无法确定归属，法律执行实际上将变得不可行。

3.4. 当前举措与展望

虽然当前的生成模型缺乏原生的归因能力，但早期的举措正在探索克服这一限制的方法[20]。其中一个例子是GEMA，即德国音乐作者和权利持有者协会。尽管他们的技术方法仍处于试点阶段且未公开披露，但他们的努力正是旨在解决大规模归因问题，提供可能允许权利持有者评估其作品是否在训练中被使用的机制[21,22]。

---

[17] 参见 Park, S. M., Georgiev, K., Ilyas, A., Leclerc, G., & Madry, A. (2023, January). TRAK: Attributing Model Behavior at Scale. In *ICML*.[18] 我们谨慎推荐修订 Abbott, R., & Rothman, E. (2023). Disrupting creativity: Copyright law in the age of generative artificial intelligence. *Fla. L. Rev.*, *75*, 1141.[19] 参见欧洲议会和理事会于2019年4月17日通过的关于数字单一市场中版权及相关权利的指令(EU) 2019/790及对指令96/9/EC和2001/29/EC的修订，欧洲共同体官方公报2019 L 130, 92[20] 参见 Stackpole, B. (2025, March 3). 为训练人工智能所用数据带来透明度. MIT SloanManagement Review. https://mitsloan.mit.edu/ideas-made-to-matter/bringing-transparency-to-data-used-to-train-artificial-intelligence[21] 参见 GEMA. (2024, November 13). Suno AI 和 OpenAI：GEMA 提起公平报酬诉讼。检索自 https://www.gema.de/en/news/ai-and-music/ai-lawsuit[22] 参见 CISAC. (2025, January 21). 要求公平报酬：GEMA 对 Suno Inc. 提起诉讼。检索自 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

GEMA's work illustrates an essential point: the difficulty of attribution should not be treated as a fixed
这并非是生成式人工智能的内在限制，而是一项有条件的技术挑战，原则上可以通过持续的研究和制度支持来解决。当前缺乏归因机制不应被解读为对创作作品无偿使用的正当理由，而应视为一个行动号召，投资于可审计且功能上可解释的生成系统的开发。

补充这些早期举措，近期研究[23,24]系统地记录了缺乏标准化数据来源实践如何导致基础模型训练流程的不透明，强调了用于追踪数据真实性、同意和许可的常用工具不足以支持大规模负责任的模型开发。他们的大规模分析强调，没有用于记录和验证数据集来源的基础设施，建立归属和问责的努力将在生成式人工智能系统中受到严重限制。

这些观点促使以下章节提出具体的技术建议，以推动GenAI中具备归属感知和透明度支持的基础设施，支持在人工内容时代实现更平衡且具法律责任的创作权生态系统。

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

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. TECHNICAL RECOMMENDATIONS TO ENHANCE TRANSPARENCY AND LEGAL COMPLIANCE IN GENAI

关键发现

- 针对可追溯性和归属挑战的技术解决方案是可行的，但需要有针对性的研究、标准化以及行业的采纳和投资。
- GenAI系统的开发者应主动承担记录数据集来源和实现模型级透明度审计的责任。
- 报酬框架应不仅反映字面复制，还应反映统计使用和对生成输出的累积影响。
- 开放标准、可审计的基础设施以及权利人、研究人员和监管者之间的结构化合作对于确保合规性和合法性至关重要。
- 欧盟处于有利位置，能够引领建立一个平衡的监管-技术生态系统，该系统既保护创作者，又促进负责任的创新。

前面的章节概述了在生成模型背景下，使归属和新颖性评估特别困难的技术基础。这些挑战并非人工智能的固有限制，而是当前设计选择和对透明度导向基础设施投资不足的结果。解决这些差距需要开发者的有意识技术承诺和更广泛的政策支持。本节提出了一套基于证据的建议，重点在于促进GenAI系统中更好的治理、问责和法律合规。

4.1. 所有利益相关者必须承担责任

解决归属和新颖性问题的责任不应仅仅由权利人或公共机构承担。生成式人工智能系统的开发者，尤其是那些在工业规模上运营的开发者，必须被期望投资于使模型可审计、可追踪并能评估其对版权数据的统计依赖性的机制。正如算法系统中个人数据的使用现在需要正式的风险评估和技术保障一样，创作作品的使用也必须受到类似的技术尽职调查框架的管理。

This includes:

- 实施系统以评估训练数据集是否包含版权内容并记录此类评估。
- 开发归属推断、影响评分或成员估计的内部能力。
- 提供技术文档和披露，支持第三方审计和法律验证。

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).

4.2. 从逐字复制到统计使用：重新思考报酬模型

当前的版权执法框架通常侧重于检测字面复制或近似衍生作品。然而，如上所述，生成系统通过学习输入数据的统计近似来运作，而非直接存储数据。因此，通知范围扩大到包括功能依赖和分布影响。 on of "use" in GenAI must be

需要技术、法律和经济方法的结合来解释：

- 模型能够整合受保护作品的统计特征，即使没有输出是逐字复制的这一事实。
- 难以识别大规模数据集中哪些部分实质性地影响生成空间的特定区域。
- 通过聚合无数微小输入（例如，数百万个短文本片段或音乐短句）所创造的累积价值，这些输入单独看似微不足道。

在此背景下，报酬方案不应仅限于字面匹配，而应探索基于统计使用的机制，例如：

- 基于数据集包含情况索引的池式版税。
- 基于影响力估算工具得出的贡献加权补偿模型。
- 数据源的公共登记册和创作者的选择性补偿系统。

4.3. 可追溯性基础设施：研究与治理的优先事项

归因和可追溯性的技术进步不仅是可能的，而且已经在进行中。诸如成员推断、数据集影响函数和语义相似性检测等新兴方法，为追踪单个数据点对模型输出的统计影响提供了日益坚实的基础。

公共投资和标准化工作应优先考虑：

- 用于审计GenAI训练管道的开源框架[25]。
- 用于衡量特定数据点影响生成样本可能性的工具。
- 数据集文档和来源标记的标准协议。
- 开发独立的测试套件以评估模型间的可追溯性和新颖性。

[25] 参见Zhong, H., Chang, J., Yang, Z., Wu, T., Mahawaga Arachchige, P. C., Pathmabandu, C., & Xue, M. (2023年4月). 生成式AI的版权保护与问责：攻击、水印和归因。载于ACM Web Conference 2023伴随会议论文集（第94-98页）。

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.

4.4. 标准化、审计与与权利人合作

建立技术标准对于确保GenAI生态系统中的互操作性和问责制至关重要。标准应当：

- 定义什么构成对训练来源的充分文档记录。
- 提供模型级归属报告的规范。
- 支持权利人和机构可用的审计机制。

除了标准和法规[26,27]之外，还必须积极鼓励学术研究人员和监管者。协作框架可以 **cooperation between GenAI developers, authors' societies,**

- 促进带注释数据集的安全共享，以测试归因方法。
- 使研究激励与监管需求保持一致。
- 确保创作者能够获得有意义且可用的工具来保护其权利。

这套建议提供了一个技术上有依据的路径，以提升生成式人工智能的透明度、归因和公平性[28]。如果没有这些措施，强大的生成式人工智能模型与原创者权利之间日益扩大的不对称，可能会破坏数字内容经济的合法性。

[26] 参见 European Commission. (2024). *Artificial Intelligence Act Recital 107*. https://ai-act-law.eu/recital/107/ [27] 参见 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] 与 - *OECD Artificial Intelligence Papers*, No. 22, OECD Publishing, Paris, https://doi.org/10.1787/2476b1a4-en.

OECD (2024), "AI, data governance and privacy: Synergies and areas of international co operation",

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.

5. CONCLUSION

当前生成式人工智能（GenAI）模型的能力基于统计近似和高维函数学习。这些系统并非通过不可解释的方式运作，而是依赖于从庞大训练数据集中得出的概率插值。虽然它们构建的内部表示，通常以复杂的超曲面形式存在，无法直接解释，但它们最终受数学原理支配，因此可以进行系统评估、改进和法规监管。

我们所描述的技术限制（见第1-3节），尤其是与记忆、新颖性和可追溯性相关的限制，并非无法解决，但它们需要专门的投资、方法创新和积极的治理。忽视这些问题将有可能破坏知识产权生态系统以及公众对负责任开发AI系统的信任。

传统的作者身份、影响力和原创性概念需要在随机生成和高维建模的背景下重新诠释。这需要跨部门的响应，整合AI研究、版权法和伦理治理的专业知识（见第4节）。

生成式AI系统为创造力和生产力开辟了变革性的可能性，但它们也带来了新的挑战。这些挑战不在于它们是否具有人类感知能力，而在于它们能够放大和重混从现有内容中学习到的统计结构。这使得它们既极具价值，又本质上与先前的人类作品纠缠在一起。

欧盟在定义生成式AI透明度、归属和责任的规范及技术基准方面具有独特的领导地位。这样做不仅将保护权利人和民主价值观，还将促进基于合法性、问责制和社会信任的创新。

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.

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.

PE 776.529 21

- 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).

- 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

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
应法律事务委员会的要求，负责自由和制度事务。

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