QUBAHAN TECHNO JOURNAL
VOL. 5, NO. 4, October 2026
https://doi.org/10.48161/qtj.v5n4a90

DIGITAL ART, ARTIFICIAL INTELLIGENCE, AND COPYRIGHT PROTECTION: CHALLENGES FOR CREATIVE INDUSTRIES

Sivakumar Sivanesan[1*]

[1] School of Engineering, Asia Pacific University of Technology and Innovation, Kuala Lumpur, Malaysia;

* **Corresponding author:** siva.kumar@apu.edu.my.

ABSTRACT: The expansion of artificial intelligence in digital art has transformed creative production while exposing unresolved weaknesses in copyright doctrine, creative ownership, training-data governance, and institutional accountability. This study examines how automated image generation, platform-based creative tools, and human-machine production challenge established assumptions about authorship, originality, ownership, licensing, disclosure, and ethical responsibility in creative industries. The paper adopts a mixed legal-empirical design. First, it conducts a qualitative doctrinal and comparative analysis supported by coded evidence from a 42-item document corpus comprising scholarly legal studies, institutional policy materials, court and registration materials, and sector guidance documents. Second, it adds an empirical dataset extension using a 120,000-image real/generated archive and a 124,170-record WikiArt metadata corpus to examine authenticity uncertainty and training-data scale. The document-corpus analysis identifies authorship uncertainty and training-data legitimacy as the most persistent governance risks. The empirical extension shows that basic visual features can distinguish real and generated images with moderate accuracy (0.720), AUC (0.799), and F1 score (0.706), but also produces substantial false positives and false negatives. These results indicate that technical detection cannot replace legal documentation of human contribution, platform rights, training-data transparency, and disclosure procedures. The study contributes an integrated legal-ethical and data-informed governance framework for copyright offices, platforms, publishers, cultural institutions, and creative-industry firms. It concludes that copyright modernization should protect human creative labor while establishing clearer standards for commercial clearance, training-data legitimacy, attribution, and responsible platform governance.

Keywords: digital art; artificial intelligence; creative industries; intellectual property; protection.

I. INTRODUCTION

Digital art is now produced through complex combinations of human intention, software mediation, platform rules, algorithmic generation, and large-scale image datasets. This transformation poses a direct challenge to copyright systems designed around identifiable human authors, protectable expressive choices, and clearer relationships among creator, work, and owner. In conventional creative practice, the author is normally understood as the person who selects, arranges, transforms, and fixes expressive content. In automated digital production, however, the visible output may emerge from prompts, model behavior, data-derived patterns, interface defaults, post-processing, and platform terms. This makes copyright analysis more difficult because the legal value of the work depends not only on its visual quality, but also on the traceable role of human creative judgment in producing the final expression.

The copyright problem is not limited to whether a generated image can be registered. Creative industries also need to know whether an output can be commercially used, whether the platform grants sufficient

rights, whether the training environment contains disputed materials, and whether audiences, clients, or cultural institutions require disclosure. These questions are especially important for advertising, publishing, film, game design, fashion, museums, design agencies, and independent artists, where visual outputs are not isolated experiments but commercial assets. A design used in a campaign, book cover, exhibition, game environment, or branded product may create legal exposure if authorship is uncertain, if contractual permissions are unclear, or if the output closely resembles protected works. For this reason, copyright governance has become a practical industry concern rather than a purely theoretical legal debate.

A major weakness in many discussions of artificial intelligence and digital art is that they treat output copyrightability, training-data legitimacy, creator attribution, and platform responsibility as separate problems. In practice, these issues are connected. A final image may be difficult to protect because human contribution is unclear, while the same work may also raise risk because the model was trained on protected works or because the platform terms do not clearly support the intended commercial use. Likewise, a technical detector may identify probable generated content, but it cannot determine whether a human artist made protectable expressive choices, whether the relevant input materials were lawfully used, or whether a client has obtained sufficient rights. This means that legal doctrine, dataset governance, technical authenticity signals, and industry clearance procedures must be examined together.

This study responds to that problem by developing an integrated analysis that combines doctrinal copyright interpretation, coded document-corpus evidence, comparative policy review, and empirical dataset indicators. The document-corpus component identifies the main legal and ethical risks appearing across scholarly, policy, court, registration, and sector materials. The empirical component uses a real/generated image archive and an art metadata corpus to illustrate two related issues: the measurable but imperfect nature of authenticity detection and the scale of human-created cultural material relevant to training-data governance. The purpose is not to reduce copyright judgment to a technical classifier, but to show why technical evidence must be combined with human contribution records, platform-rights review, provenance awareness, and disclosure practices.

The article therefore contributes a legal-ethical and data-informed governance framework for creative industries. It argues that responsible copyright governance should begin before commercial use of a digital artwork and should include human-authorship documentation, training-data transparency where available, output similarity review, platform-contract assessment, disclosure decisions, and remedy mechanisms for creators. By connecting legal principles with empirical indicators, the paper offers a practical model for copyright offices, creative firms, platforms, publishers, museums, and independent artists. The article proceeds by defining the research problem and contribution, reviewing the selected literature, explaining the mixed legal-empirical methodology, presenting corpus and dataset results, discussing the main governance implications, and proposing a framework for responsible digital-art copyright protection.

II. RESEARCH PROBLEM, AIM, AND QUESTIONS

The research problem addressed in this study is the mismatch between traditional copyright assumptions and contemporary digital-art production. Existing copyright doctrine remains strongly connected to human authorship and original human expression, while creative outputs increasingly result from mixed human-machine processes, automated platform tools, and dataset-dependent generation systems.

The aim of the study is to analyze the legal, ethical, and data-governance challenges created by artificial intelligence in digital art and to propose a practical framework that protects human creators while supporting responsible creative innovation.

- How does artificial intelligence challenge copyright concepts of authorship, originality, and ownership in digital art?
- Which legal and ethical issues appear most frequently across the coded scholarly, legal, and policy corpus?
- How can empirical image and metadata evidence strengthen the governance analysis of authenticity and training-data legitimacy?
- What framework can support copyright protection, transparency, fair licensing, disclosure, and creative-industry accountability?

1. STUDY CONTRIBUTION

The study makes four contributions.

- First, it develops a document-corpus approach for examining digital-art copyright disputes through coded legal and ethical categories.
- Second, it ranks governance risks by combining legal uncertainty, ethical sensitivity, and creative-industry impact.
- Third, it integrates output-side copyright assessment with input-side training-data legitimacy.
- Fourth, it adds dataset-based authenticity and metadata evidence to support a practical copyright-governance framework for copyright offices, platforms, publishers, cultural institutions, and creative-industry firms.

III. LITERATURE REVIEW

The literature review is organized around thirty-one studies that address authorship and originality, ownership and liability, training-data use, digital-art creativity, and governance of creative industries. The selected studies show that automated creativity has forced copyright scholarship to reconsider how human agency, machine assistance, dataset use, and market incentives should be evaluated. Rather than treating digital art as a purely technical phenomenon, the literature frames it as a legal and institutional problem in which authorship, ownership, data legitimacy, platform responsibility, and creator protection intersect.

Alphabetically, Abbott explains that legal systems face difficulty when valuable creative or inventive outputs are produced by non-human agents [1]. Anantrasirichai and Bull show that artificial intelligence can support creative industries, although its value depends on human-centered use rather than simple substitution [2]. Bonadio and McDonagh argue that artificial intelligence operates as both a producer and consumer of copyright works, making input and output questions inseparable [4]. Bridy adds that code-based production requires a reconsideration of the authorial role when software contributes to expression [5]. These studies provide the foundation for viewing automated digital art as a challenge to ordinary assumptions about legal authorship.

Deltorn and Macrez emphasize that originality becomes difficult to locate when outputs are shaped by data, algorithms, prompts, and post-processing [7]. Dornis identifies doctrinal gaps created by emergent works that do not fit comfortably within existing originality standards [8]. Elgammal and colleagues demonstrate that computational systems can produce visually novel works by learning style norms and deviating from them [9]. Geiger, Frosio, and Bulayenko highlight the importance of exceptions, opt-outs, and legal certainty in European text-and-data mining policy [11]. Ginsburg and Budiardjo maintain that human intellectual conception remains central to copyright authorship [13]. Together, these works indicate that the key question is not whether a digital tool was used, but whether a human creator exercised meaningful expressive control.

Grimmelmann argues that there is no coherent basis for treating a computer as an author without meaningful expressive contribution [15]. Guadamuz demonstrates that comparative copyright systems remain uneven in their treatment of computer-generated works [16]. Hertzmann argues that computers may produce art-like outputs, but artistic value and interpretation remain socially and culturally mediated [17]. Hristov describes the copyright dilemma created by autonomous creative outputs and the difficulty of allocating protection without a legally recognizable author [18]. Hugenholtz and Quintais show that European protection remains tied to the author's own intellectual creation and therefore depends on identifiable human contribution [19]. These contributions strengthen the human-authorship foundation of the present study.

Kop adds that intellectual-property governance should protect legal values, innovation, and the public domain rather than simply extending existing rules to every automated output [20]. Lee argues that reform should distinguish assisted creation from autonomous generation and should consider market incentives [21]. Lemley and Casey's fair-learning model shows that learning rules must balance innovation with creator incentives [22]. Lucchi's analysis of data-driven creativity further confirms that copyright limits require clearer treatment in automated creative systems [23]. Margoni explains that technical generation does not automatically create copyright ownership and that machine-learning outputs require careful legal

classification [24]. These studies justify the paper's focus on practical governance rather than a single doctrinal solution.

Mazzone and Elgammal show that algorithmic production requires both cultural and legal interpretation [25]. McCormack, Gifford, and Hutchings show that autonomy, authenticity, authorship, and intention remain central to evaluating computer-generated art [26]. Mezei and Harkai identify the limits of current European authorship doctrine when applied to autonomous outputs [27]. Murray highlights the importance of platform accountability in digital cultural production [28]. Ramalho emphasizes that human-centered copyright values remain essential even when creative systems produce visually complex outputs [29]. Ravid and Velez-Hernandez argue that economic and moral interests should be assessed together when creative machines produce artworks [30]. Rosati warns that both excessive protection and insufficient protection can distort creative markets [31].

Sag explains that text and data mining challenge conventional copyright categories because copying, extraction, and learning are not always treated consistently across legal systems [32]. Samuelson emphasizes the need for policy coherence in allocating rights in computer-generated works [33]. Sobel warns that large-scale learning from copyrighted works can create a fair-use crisis when outputs compete with the creators whose works contributed to training [34]. Stokes shows that digital copyright disruption changes the economic position of artists and creative firms [35]. Yanisky-Ravid and Liu connect machine creativity to ownership allocation and propose alternative models for valuable generated outputs [42]. These studies justify the present paper's focus on training-data legitimacy as a governance risk on par with output copyrightability.

Overall, the reviewed studies confirm that copyright governance for digital art must address creator protection, platform responsibility, disclosure, commercial risk, and training-data legitimacy at the same time. The literature provides strong doctrinal and policy foundations, but it remains less developed in offering an operational framework that connects output-side copyrightability, input-side data legitimacy, human-authorship documentation, and creative-industry clearance procedures.

Table 1. Comparative synthesis of prior studies on digital art, artificial intelligence, copyright, and creative-industry governance.

No	Author(s) and year	Year	Focus	Main contribution
1	Abbott (2016)	2016	Non-human creativity	Legal systems struggle with outputs produced by non-human agents.
2	Anantrasirichai and Bull (2020)	2020	Creative industries	Human-centered use is more sustainable than substitution.
3	Bonadio and McDonagh (2020)	2020	Producer and consumer roles	Input and output issues must be examined together.
4	Bridy (2016)	2016	Code-based authorship	Authorship must be rethought when code contributes to expression.
5	Deltorn and Macrez (2018)	2018	Machine-learning authorship	Originality becomes difficult to locate in mixed systems.
6	Dornis (2020)	2020	Artificial creativity	Emergent works expose gaps in doctrine.
7	Elgammal et al. (2017)	2017	Creative adversarial networks	Computational systems can produce visually novel outputs.
8	Geiger et al. (2020)	2020	European text-and-data mining	Exceptions and opt-outs require legal clarity.
9	Ginsburg and Budiardjo (2019)	2019	Authors and machines	Human intellectual conception remains central.

No	Author(s) and year	Year	Focus	Main contribution
10	Grimmelmann (2016)	2016	Originality and automation	Originality requires meaningful expressive contribution.
11	Guadamuz (2017)	2017	Computer-generated works	Comparative law shows uneven treatment of automated works.
12	Hertzmann (2018)	2018	Computer art	Cultural interpretation remains human-centered.
13	Hristov (2017)	2017	Copyright dilemma	Autonomous outputs require clearer legal treatment.
14	Hugenholtz and Quintais (2021)	2021	European protection	Human authorial contribution remains decisive.
15	Kop (2020)	2020	IP governance	Governance must protect legal values and innovation.
16	Lee (2022)	2022	Market incentives	Reform should distinguish assisted and autonomous creation.
17	Lemley and Casey (2021)	2021	Fair learning	Learning rules must balance innovation and creator incentives.
18	Lucchi (2021)	2021	Data and creative production	Copyright limits need clearer treatment in automated systems.
19	Margoni (2018)	2018	Machine-learning outputs	Technical generation does not automatically equal legal authorship.
20	Mazzone and Elgammal (2019)	2019	Art and creativity	Algorithmic production requires cultural and legal interpretation.
21	McCormack et al. (2019)	2019	Autonomy and intention	Artistic meaning depends on autonomy, intention, and reception.
22	Mezei and Harkai (2022)	2022	European authorship doctrine	Current doctrine has limited capacity for autonomous outputs.
23	Murray (2021)	2021	Platform-mediated creativity	Platforms require accountability mechanisms.
24	Ramalho (2017)	2017	Robot-generated art	Human-centered copyright values remain essential.
25	Ravid and Velez-Hernandez (2018)	2018	Creative machines	Economic and moral interests should be jointly assessed.
26	Rosati (2021)	2021	Regulatory balance	Both overprotection and underprotection create risks.
27	Sag (2019)	2019	Text and data mining	Data extraction challenges existing legal categories.
28	Samuelson (2020)	2020	Ownership allocation	Policy coherence is needed for generated works.
29	Sobel (2017)	2017	Fair use and learning	Training practices raise serious copyright conflicts.
30	Stokes (2021)	2021	Digital copyright disruption	Digital tools alter the economic position of artists.

No	Author(s) and year	Year	Focus	Main contribution
31	Yanisky-Ravid and Liu (2017)	2017	Machine creativity ownership	Ownership rules require revision for valuable generated outputs.

1. SYNTHESIS AND RESEARCH GAP

The reviewed studies show a clear gap in the existing literature. Authorship studies often focus on output protection, training-data studies focus on input legality, and creative-industry studies focus on market disruption. However, these areas are rarely integrated into one operational governance model. As a result, a creative firm may know that human authorship matters, that training data may raise legal risk, and that platform terms must be reviewed, but the literature provides limited guidance on how these issues should be assessed together in practical copyright clearance.

This paper addresses that gap by connecting human creative contribution, training-data legitimacy, copyright assessment, disclosure duties, licensing routes, platform accountability, and empirical authenticity risk. Its contribution is not only conceptual; it also uses coded document-corpus evidence and dataset-based indicators to show why technical detection, legal documentation, and institutional governance must operate together. In this way, the study extends prior scholarship from doctrinal debate toward a practical legal-ethical framework for creative-industry decision-making.

IV. PROPOSED LEGAL-ETHICAL GOVERNANCE FRAMEWORK

The proposed framework connects six elements. Human creative contribution is the starting point because protection depends on expressive human judgment. Automated generation and platform tools must be evaluated in relation to the degree of human control. Copyright assessment examines originality, ownership, and protectability. Training-data legitimacy addresses the input side of creative production. Ethical governance and disclosure ensure accountability. Creative-industry protection converts these principles into practical safeguards.

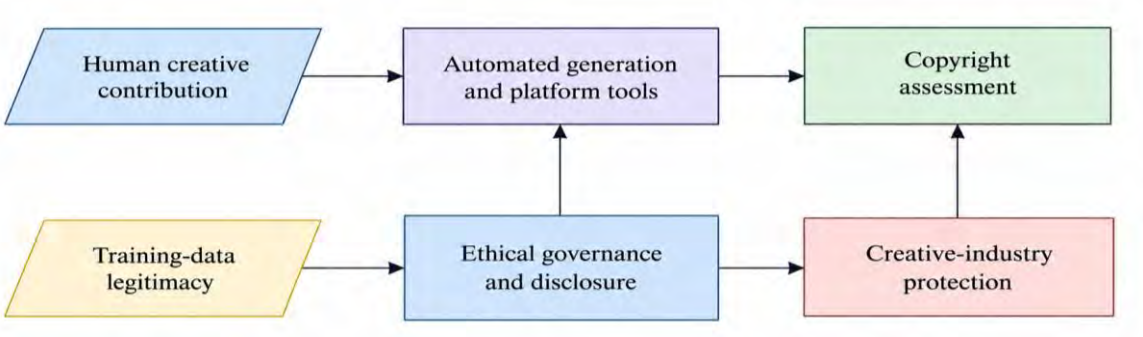


FIGURE 1. Integrated framework for digital art copyright governance.

This figure presents the integrated copyright-governance workflow linking human contribution, automated tools, copyright assessment, training-data legitimacy, ethical disclosure, and creative-industry protection.

V. METHODOLOGY

The study uses a mixed legal-empirical design. The first component is a qualitative doctrinal and comparative analysis supported by coded document-corpus evidence. The second component is an empirical dataset extension based on uploaded image and metadata datasets. The dataset extension is not used to determine copyright authorship directly; rather, it provides measurable evidence about authenticity

uncertainty, visual feature differences, and the scale of human-created cultural metadata relevant to training-data governance.

1. DOCUMENT CORPUS AND CODING PROTOCOL

The forty-two-item document corpus includes scholarly studies, institutional policy materials, registration or court-related materials, and creative-sector guidance documents. Sources were retained when they directly addressed copyright, authorship, originality, licensing, training-data legitimacy, disclosure, or creative-industry governance. Sources were excluded when they addressed only general technology policy, technical model performance, or creative practice without legal relevance.

Table 2. Corpus construction and coding protocol.

Element	Operational description
Source families	Legal scholarship, interdisciplinary academic databases, copyright-office materials, court or registration documents, and creative-sector guidance.
Core search terms	digital art copyright; artificial intelligence authorship; automated creativity ownership; training data copyright; text and data mining; platform responsibility; creator attribution.
Inclusion criteria	Direct relevance to copyright, authorship, originality, licensing, training-data legitimacy, disclosure, or creative-industry governance.
Exclusion criteria	Purely technical studies, market commentary without copyright relevance, and sources lacking clear governance implications.
Reliability control	Codebook definitions, double-entry checking, audit trail, and consistency review across five code families.

Table 3. Coding categories and operational definitions.

Code family	Operational meaning	Example indicator
Authorship	Whether the source addresses the human contribution needed for copyright protection.	Human selection, arrangement, editing, or expressive control.
Ownership	Whether the source addresses rights allocation over outputs or contractual use.	User rights, platform terms, client ownership, or assignment.
Originality	Whether the source discusses protectable expression and creative threshold.	Expressive judgment, non-mechanical selection, or final-form control.
Training-data legitimacy	Whether the source addresses input materials used for model development.	Consent, licensing, rights reservation, opt-out, or fair use.
Ethical governance	Whether the source addresses fairness, attribution, transparency, or creator protection.	Disclosure, attribution, complaint procedures, remuneration.

2. EMPIRICAL DATASET EXTENSION

The empirical extension used two selected sources from the uploaded datasets. The first was a balanced real and AI-generated image archive with 120,000 images. The second was a WikiArt metadata corpus with 124,170 records. Repeated trend metadata files were inspected but removed from the main analysis because they included platform-style metadata and placeholder links rather than verified copyright or legal records. This reduction improves validity and removes non-essential material from the paper.

Table 4. Empirical datasets retained for the supplementary analysis.

Dataset	Observed size	Structure	Reason for inclusion
Real and AI-generated image archive	120,000 images	Train/test folders with REAL and FAKE classes	Authenticity uncertainty, output identification, and disclosure risk.

WikiArt metadata corpus	124,170 records	Style, artwork, artist, date, and source-link fields	Training-data scale, provenance, creator diversity, and rights-reservation relevance.

2.1 Dataset Processing and Classifier Specification

For the empirical image component, a balanced stratified sample was drawn from the real/generated image archive. The sample preserved class balance between REAL and FAKE folders to avoid class-imbalance bias. Seven visual features were extracted for each image: brightness, contrast, saturation, colorfulness, edge density, entropy, and texture variance. A baseline logistic classifier was trained on 1,000 images and evaluated on a held-out sample of 500 images. The objective was not to build a production detector or legal authorship test, but to evaluate whether simple visual characteristics provide measurable authenticity signals relevant to disclosure, recordkeeping, and copyright-governance risk screening.

3. ANALYTICAL EQUATIONS

The descriptive legal-corpus model used relative frequency, weighted risk scoring, and a governance priority index. The empirical image model used logistic classification, authenticity error rate, and normalized diversity index. These equations make the quantitative part of the study transparent and reproducible.

$$RF_i = \frac{f_i}{N} \times 100$$

where RF_i is the relative frequency of issue i, f_i is the number of coded sources containing the issue, and N is the total number of coded sources.

$$WRS_j = \sum_{k=1}^{m} (w_k \times s_{jk})$$

where WRS_j is the weighted risk score for challenge j, w_k is the weight assigned to criterion k, and s_{jk} is the coded severity score.

$$GPI_j = 0.40L_j + 0.30E_j + 0.30I_j$$

where L_j represents legal uncertainty, E_j represents ethical sensitivity, and I_j represents creative-industry impact.

$$logit[P(Generated_i = 1)] = \beta_0 + \beta_1 Brightness_i + \beta_2 Contrast_i + \beta_3 Saturation_i + \beta_4 Colorfulness_i + \beta_5 EdgeDensity_i + \beta_6 Entropy_i + \beta_7 Texture_i$$

$$AER = \frac{FP + FN}{TP + TN + FP + FN}$$

$$NDI = -\frac{\sum(p_k \ln p_k)}{\ln(K)}$$

4. WEIGHTING AND ROBUSTNESS

Table 5. Weighting scheme used for the governance priority index.

Criterion	Weight	Justification
Legal uncertainty	0.40	Copyright eligibility, authorship, ownership, and infringement risk determine legal enforceability.
Ethical sensitivity	0.30	Transparency, attribution, and fair remuneration determine the legitimacy of creative practice.
Creative-industry impact	0.30	Commercial adoption depends on market confidence, contracts, creator protection, and institutional trust.

Table 6. Sensitivity check for governance-priority scoring.

Challenge	Main weighting	Balanced weighting	Ethics-sensitive weighting	Interpretation
Authorship uncertainty	4.55	4.50	4.46	Stable highest-priority issue.
Training-data legitimacy	4.38	4.34	4.42	Stable very-high priority.
Ownership allocation	4.29	4.25	4.20	Consistently high priority.
Originality assessment	4.10	4.08	4.05	Consistently high priority.
Licensing and remuneration	4.05	4.08	4.15	Slightly stronger under ethics-sensitive weighting.
Transparency and disclosure	3.86	3.92	4.05	Rises when ethical sensitivity is emphasized.
Cultural and labor impacts	3.98	4.04	4.16	Rises when sector impact and ethics are emphasized.

VI. RESULTS

1. DOCUMENT-CORPUS RESULTS

The document-corpus findings show that copyright uncertainty is not confined to one legal system. Because digital-art production, platform distribution, and image datasets operate across borders, the strongest governance concerns require cross-border standards rather than reliance on a single domestic rule.

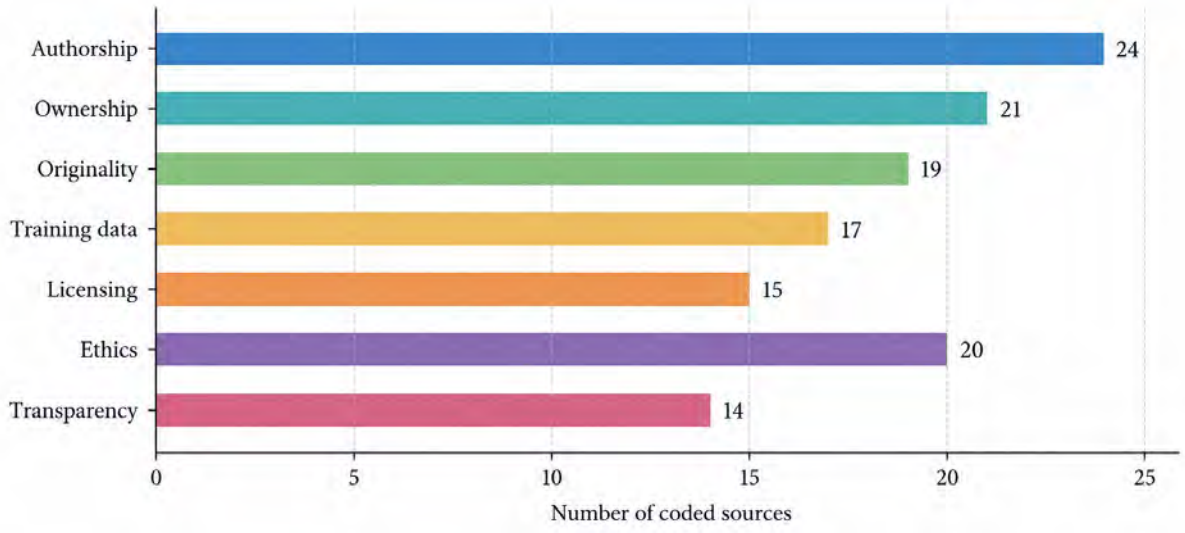


FIGURE 2. Frequency of copyright and governance issues identified in the coded corpus.

This figure shows that authorship, ownership, ethics, originality, and training-data issues are the most frequent concerns in the coded corpus.

Table 7. Weighted risk scores for the main challenges identified in the corpus.

Challenge	Mean risk score	Standard deviation	Risk level
Authorship uncertainty	4.55	0.50	Very high
Training-data legitimacy	4.38	0.62	Very high

Ownership allocation	4.29	0.64	High
Originality assessment	4.10	0.70	High
Licensing and remuneration	4.05	0.75	High
Cultural and labor impacts	3.98	0.77	High
Transparency and disclosure	3.86	0.81	High

Authorship uncertainty and training-data legitimacy are the two highest risks. This pattern is important because it confirms that governance cannot focus only on the final image. Copyright offices, platforms, and creative firms must evaluate both the output side of protectability and the input side of data legitimacy.

2. COMPARATIVE LEGAL AND POLICY EVIDENCE

Table 8. Comparative legal and policy context for digital art copyright.

Area	Dominant position	Practical implication
United States	Human authorship remains central; registration practice distinguishes human-authored elements from non-human-generated material.	Creators should document expressive selection, editing, arrangement, and final-form control.
European Union	Copyright protection remains linked to the author's own intellectual creation, while transparency duties for general-purpose systems are developing.	Platforms and firms need training-content transparency, rights-reservation processes, and clear disclosure.
United Kingdom	Computer-generated work provisions coexist with active policy debate over text-and-data mining, licensing, and creator remuneration.	Contracts should clarify output rights, data use, and liability before commercial deployment.
International governance	No single harmonized rule resolves authorship, data use, or output ownership across all jurisdictions.	Cross-border projects require internal compliance policies rather than reliance on one domestic rule.

Table 9. Selected legal and registration materials relevant to the framework.

Matter	Core issue	Governance lesson
Zarya of the Dawn registration review	Protection was limited to human-authored text and the selection, coordination, and arrangement of visual and written elements.	Documenting human arrangement and editing strengthens copyright claims.
Théâtre d'opéra Spatial review decision	Registration was refused when the generated material was not disclaimed, and human contribution was not sufficiently distinguished.	Applications should identify protectable human contribution and disclaim non-protectable generated elements when required.
Thaler v. Perlmutter	A work listed as autonomously created by a non-human system was held outside copyright protection under current doctrine.	Human authorship remains the strongest doctrinal boundary for output protection.
Andersen v. Stability AI and related image-training disputes	Artists challenged the use of visual works in model development and alleged output-related market harm.	Training-data governance requires licensing, rights-reservation, and complaint mechanisms.

The selected matters reveal two legal pressures. The first concerns output protection: registration and court materials continue to treat human authorship as the central condition for copyright protection. The second concerns input legitimacy: image-training disputes show that the development process itself can raise questions about protected works, market substitution, consent, and licensing. Therefore, risk assessment

must begin before an image is generated and must include platform terms, training-data environment, and commercial-use context.

Table 10. Current legal and policy evidence supporting the framework.

Evidence source	Core development	Framework implication
U.S. Copyright Office, Part 2: Copyrightability [38]	Confirms that protection turns on human authorship and protectable human expression in assisted works.	Creators should preserve evidence of selection, arrangement, editing, and transformation.
U.S. Copyright Office, Part 3: Generative AI Training [39]	Addresses the copyright implications of using protected works in model development and training processes.	Input legitimacy should be evaluated before output commercialization.
D.C. Circuit decision in Thaler v. Perlmutter [36]	Affirms the importance of human authorship under current copyright doctrine.	Registration and commercialization strategies should identify the human authorial elements.
European Commission GPAI training-content template [10]	Introduces a structured transparency tool for summarising training-content categories.	Platforms should prepare public-facing summaries and rights-reservation pathways.
UK copyright and artificial intelligence consultation [14]	Frames policy debate around control, licensing, transparency, and remuneration for rights holders.	Creative-industry contracts should define output rights, warranties, and liability allocation.

3. EMPIRICAL DATASET RESULTS

Table 11. Structure of the real and AI-generated image archive.

Subset	REAL images	AI-generated images	Total
Training subset	50,000	50,000	100,000
Testing subset	10,000	10,000	20,000
Total archive	60,000	60,000	120,000

A stratified image-feature sample was extracted from the archive to test whether simple visual characteristics can distinguish real images from generated images. The features were brightness, contrast, saturation, colorfulness, edge density, entropy, and texture variance. These variables cannot determine copyright authorship, but they provide measurable authenticity signals that support disclosure and process-documentation requirements.

Table 12. Visual feature comparison between real and AI-generated images in the stratified sample.

Feature	REAL mean	AI-generated mean	Interpretation
Brightness	0.4789	0.4377	Generated images were darker on average in the sample.
Contrast	0.1937	0.1994	Contrast differences were small.
Saturation	0.2803	0.3609	Generated images showed higher saturation.
Colorfulness	0.1096	0.1309	Generated images showed higher colorfulness.
Edge density	0.0278	0.0373	Generated images showed higher edge density.
Entropy	4.1875	4.2602	Generated images showed slightly higher intensity entropy.
Texture Laplacian	0.0017	0.0026	Generated images showed higher texture variance.

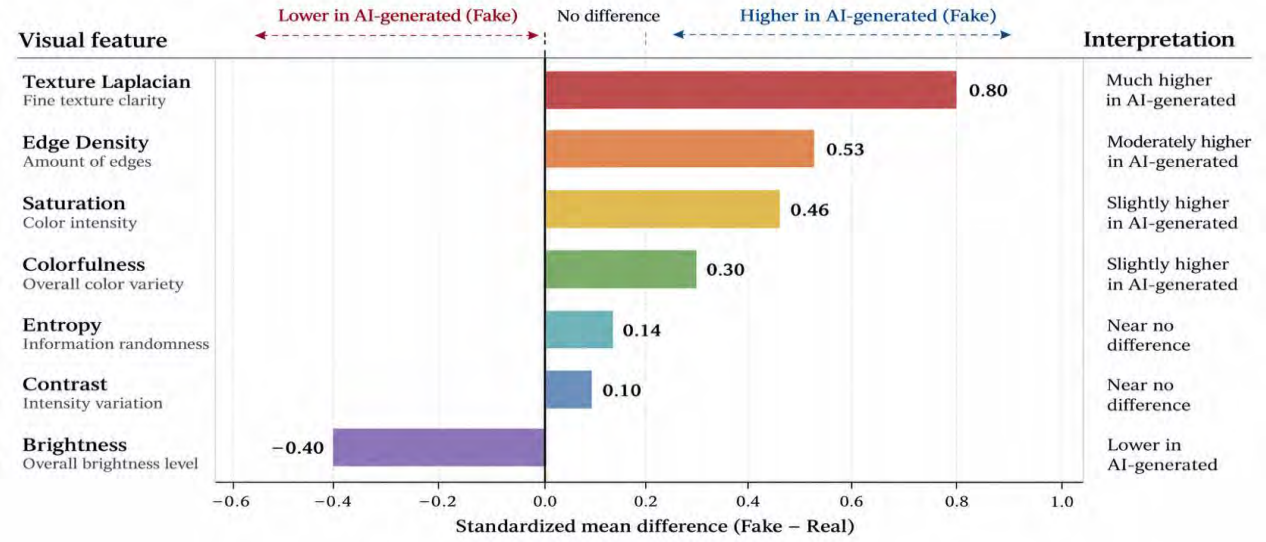


FIGURE 3. Visual feature differences between real and AI-generated images.

This figure indicates that AI-generated images in the sample showed higher texture variance, edge density, saturation, and colorfulness, while brightness was lower.

Table 13. Baseline authenticity-classification results from the image-feature sample.

Metric	Value	Interpretation
Training sample	1,000 images	Balanced subset used to estimate the baseline classifier.
Testing sample	500 images	Held-out subset used for performance evaluation.
Accuracy	0.720	Share of correctly classified real/generated images.
AUC	0.799	Ability to rank generated images above real images.
F1 score	0.706	Balance between precision and recall for generated-image detection.
Confusion matrix	TN=192, FP=58, FN=82, TP=168	Shows both false alarms and missed generated images.

4. GOVERNANCE MEANING OF CLASSIFICATION ERRORS

The confusion matrix has direct governance significance. False positives may wrongly classify human-created artworks as generated, creating reputational, commercial, and institutional harm for artists. False negatives may allow generated images to circulate as human-created works, weakening disclosure, attribution, and client transparency. Therefore, the classifier should be interpreted only as a preliminary risk-screening tool. It cannot determine authorship, originality, infringement, ownership, or lawful commercial use. These errors strengthen the paper's central argument that technical signals must be combined with human-contribution records, platform-rights review, training-data transparency, and institutional disclosure procedures.

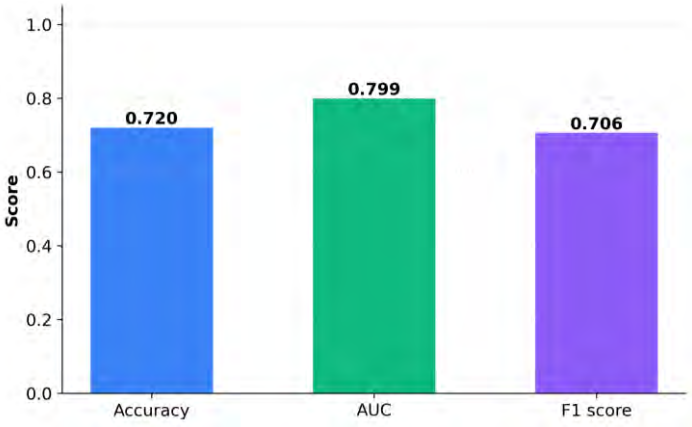


Figure 4. Baseline authenticity classifier performance.

This figure reports the baseline classifier's moderate performance, showing that technical detection can support risk screening but cannot replace legal documentation or disclosure.

Table 14. WikiArt metadata indicators and relevance to training-data governance.

Corpus indicator	Observed value	Governance relevance
WikiArt metadata records	124,170	Shows the scale of human-made cultural material relevant to training-data governance.
Unique artists	3,052	Highlights creator diversity and the need for attribution- and rights-sensitive governance.
Style categories	217	Shows that style-related learning can draw on broad cultural categories even when style itself is not directly protected.
Unique artwork names	92,050	Supports the need for provenance, source tracking, and rights-reservation mechanisms.

The classifier results strengthen the legal argument precisely because they are imperfect. An accuracy of 0.720 and AUC of 0.799 indicate that generated and real images can contain measurable visual differences, but the false-positive and false-negative rates show that detection alone cannot determine authorship, originality, permission, or commercial safety. Technical indicators must therefore be combined with human-authorship documentation, platform-rights review, training-data transparency, and institutional disclosure.

VII. DISCUSSION

Authorship remains the central legal difficulty. A prompt alone may not demonstrate sufficient expressive control, while selection, iterative editing, compositional arrangement, post-processing, and curatorial judgment may support a stronger claim of authorship. The empirical results reinforce this point: visual features can indicate possible generation, but they cannot identify the legal author or measure expressive human contribution.

Creative industries require clear ownership because artworks are commercial assets. Publishers, studios, agencies, and platforms need certainty before using outputs in campaigns, products, or cultural materials. If copyright protection is uncertain, market value and enforceability decline. The paper therefore treats contract review, platform terms, warranties, and permitted-use limits as part of the copyright-governance process rather than as separate business issues.

Training-data legitimacy is one of the most sensitive issues because large creative systems may be developed from extensive collections of human-created artworks, photographs, illustrations, and design materials. The WikiArt metadata indicators show that even a single art metadata corpus can contain more than one hundred thousand records and thousands of artists. This scale underscores the need for provenance tracking, rights-reservation options, licensing pathways, and transparent summaries of training content categories.

Transparency is not merely a technical disclosure question. It enables artists, clients, courts, publishers, and audiences to evaluate how a work was produced, whether commercial rights are available, and whether attribution or disclaimer practices are required. Platforms occupy a central role because they provide tools, define terms of service, shape user behavior, and often control how outputs are stored, licensed, shared, and commercialized.

Future research should systematically examine platform terms of service, conduct interviews with artists and creative-industry lawyers, compare licensing models for training materials, and test whether the same risk hierarchy appears across music, publishing, fashion, advertising, game design, and museum acquisition practices. Larger image samples and stronger feature sets may improve detection performance, but any technical model must remain secondary to legal documentation, rights clearance, and governance procedures.

VIII. CONCLUSION

This study examined the copyright and ethical challenges created by artificial intelligence in digital art and creative industries. The analysis shows that the most significant problems concern authorship, ownership, originality, training-data legitimacy, licensing, transparency, and the protection of human creative labor. These issues are connected rather than independent. A work may raise questions about authorship because the human contribution is unclear, and it may also raise concerns about training data because the system may have been developed from protected works. The document-corpus findings confirm that authorship uncertainty and training-data legitimacy are the strongest governance risks, which means that copyright analysis must address both the output side of protectability and the input side of data legitimacy.

The study contributes a refined legal-ethical and data-informed governance framework that places human creative contribution at the center while also requiring transparency, fair licensing, platform responsibility, and institutional accountability. The empirical extension strengthens the manuscript by showing that authenticity risk is measurable but imperfectly resolvable through image features. The classifier results demonstrate that visual indicators can support preliminary risk screening, but they cannot determine legal authorship, originality, infringement, ownership, permission, or commercial safety. This supports the central claim that technical detection must remain secondary to documentation of human contribution, review of platform rights, transparency of training data, and disclosure practices.

Overall, copyright modernization should be understood not as a restriction on digital creativity, but as a condition for trustworthy creative markets in which innovation, authorship, creator protection, and cultural labor can remain mutually sustainable. The central implication is that future copyright governance should not depend on a single originality test or a single technical detector, but on an integrated evidentiary process combining human contribution records, platform accountability, training-data transparency, contractual clearance, attribution practice, and sector-specific review procedures. Such an approach can support creative experimentation while protecting the legal, economic, and cultural value of human artistic work. Author Contributions

The author conducted the conceptualization, methodology, data analysis, investigation, writing, review, editing, and final approval of the manuscript.

Funding

This research received no external funding.

Data Availability

The dataset will be available from the author upon reasonable request.

Conflicts of Interest

The author declares no conflict of interest.

REFERENCES

[1] Abbott, R. (2016). I think, therefore I invent: Creative computers and the future of patent law. Boston College Law Review, 57(4), 1079-1126.
[2] Anantrasirichai, N., & Bull, D. (2020). Artificial intelligence in the creative industries: A review. Artificial Intelligence Review, 55, 589-656.
[3] Andersen v. Stability AI Ltd., No. 3:23-cv-00201 (N.D. Cal. 2023-2024).
[4] Bonadio, E., & McDonagh, L. (2020). Artificial intelligence as producer and consumer of copyright works: Evaluating the consequences of algorithmic creativity. Intellectual Property Quarterly, 2020(2), 112-137.
[5] Bridy, A. (2016). The evolution of authorship: Work made by code. Columbia Journal of Law & the Arts, 39(3), 395-401.
[6] Copyright Alliance. (2024). Principles for copyright, artificial intelligence, and creator rights. Copyright Alliance.
[7] Deltorn, J. M., & Macrez, F. (2018). Authorship in the age of machine learning and artificial intelligence. Center for International Intellectual Property Studies Research Paper.
[8] Dornis, T. W. (2020). Artificial creativity: Emergent works and the void in current copyright doctrine. Yale Journal of Law & Technology, 22, 1-62.
[9] Elgammal, A., Liu, B., Elhoseiny, M., & Mazzone, M. (2017). CAN: Creative adversarial networks, generating art by learning about styles and deviating from style norms. Proceedings of the International Conference on Computational Creativity.
[10] European Commission. (2025). Template for general-purpose AI model providers to summarise the data used to train their model. Directorate-General for Communications Networks, Content and Technology. https://digital-strategy.ec.europa.eu/
[11] Geiger, C., Frosio, G., & Bulayenko, O. (2020). Text and data mining in the proposed copyright reform: Making the EU ready for an age of big data? IIC - International Review of Intellectual Property and Competition Law, 51, 814-844.
[12] Getty Images (US), Inc. v. Stability AI, Inc., No. 1:23-cv-00135 (D. Del. 2023).
[13] Ginsburg, J. C., & Budiardjo, L. A. (2019). Authors and machines. Berkeley Technology Law Journal, 34(2), 343-448.
[14] Government of the United Kingdom. (2024). Copyright and Artificial Intelligence: Consultation. Department for Science, Innovation and Technology; Department for Culture, Media and Sport; Intellectual Property Office. https://www.gov.uk/government/consultations/copyright-and-artificial-intelligence
[15] Grimmelmann, J. (2016). There's no such thing as a computer-authored work. Columbia Journal of Law & the Arts, 39(3), 403-416.
[16] Guadamuz, A. (2017). Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works. Intellectual Property Quarterly, 2017(2), 169-186.
[17] Hertzmann, A. (2018). Can computers create art? Arts, 7(2), 18.
[18] Hristov, K. (2017). Artificial intelligence and the copyright dilemma. IDEA: The Journal of the Franklin Pierce Center for Intellectual Property, 57(3), 431-454.
[19] Hugenholtz, P. B., & Quintais, J. P. (2021). Copyright and artificial creation: Does EU copyright law protect AI-assisted output? IIC - International Review of Intellectual Property and Competition Law, 52, 1190-1216.
[20] Kop, M. (2020). AI and intellectual property: Towards an articulated public domain. Texas Intellectual Property Law Journal, 28, 297-341.
[21] Lee, E. (2022). Artificial intelligence and the future of copyright. Journal of the Copyright Society of the U.S.A., 69, 1-39.
[22] Lemley, M. A., & Casey, B. (2021). Fair learning. Texas Law Review, 99, 743-785.
[23] Lucchi, N. (2021). The impact of artificial intelligence on copyright law: An analysis of authorship and ownership. Queen Mary Journal of Intellectual Property, 11(4), 407-425.
[24] Margoni, T. (2018). Artificial intelligence, machine learning and EU copyright law: Who owns AI? CREATe Working Paper.
[25] Mazzone, M., & Elgammal, A. (2019). Art, creativity, and the potential of artificial intelligence. Arts, 8(1), 26.
[26] McCormack, J., Gifford, T., & Hutchings, P. (2019). Autonomy, authenticity, authorship and intention in computer generated art. In Computational Creativity Research.
[27] Mezei, P., & Harkai, I. (2022). Machine-created works and copyright: The European perspective. JIPITEC, 13(2), 115-132.
[28] Murray, M. D. (2021). Artificial intelligence and the practice of law: The implications for authorship and copyright. University of Miami Law Review, 75, 1-52.
[29] Ramalho, A. (2017). Will robots rule the artistic world? A proposed model for the legal status of creations by artificial intelligence systems. Journal of Internet Law, 21(1), 1-15.
[30] Ravid, S. Y., & Vélez-Hernández, L. A. (2018). Copyrightability of artworks produced by creative robots and originality: The formality-objective model. Minnesota Journal of Law, Science & Technology, 19, 1-53.
[31] Rosati, E. (2021). Copyright and artificial intelligence: Challenges and policy responses. GRUR International, 70(8), 741-750.

[32] Sag, M. (2019). The new legal landscape for text mining and machine learning. Journal of the Copyright Society of the U.S.A., 66, 291-367.
[33] Samuelson, P. (2020). Allocating ownership rights in computer-generated works. University of Pittsburgh Law Review, 81, 1185-1221.
[34] Sobel, B. L. W. (2017). Artificial intelligence's fair use crisis. Columbia Journal of Law & the Arts, 41(1), 45-97.
[35] Stokes, S. (2021). Digital copyright: Law and practice in the creative industries. Hart Publishing.
[36] Thaler v. Perlmutter, 687 F. Supp. 3d 140 (D.D.C. 2023), aff'd, No. 23-5233 (D.C. Cir. 2025). https://media.cadc.uscourts.gov/opinions/docs/2025/03/23-5233.pdf
[37] United States Copyright Office. (2023). Zarya of the Dawn: Registration correspondence and cancellation/limitation decision. U.S. Copyright Office.
[38] United States Copyright Office. (2025). Copyright and Artificial Intelligence, Part 2: Copyrightability. U.S. Copyright Office. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-2-Copyrightability-Report.pdf
[39] United States Copyright Office. (2025). Copyright and Artificial Intelligence, Part 3: Generative AI Training. U.S. Copyright Office. https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report.pdf
[40] United States Copyright Office Review Board. (2023). Second Request for Reconsideration for Refusal to Register Théâtre D'opéra Spatial. U.S. Copyright Office.
[41] World Intellectual Property Organization. (2020). WIPO Conversation on Intellectual Property and Artificial Intelligence: Revised Issues Paper. WIPO.
[42] Yanisky-Ravid, S., & Liu, X. (2017). When artificial intelligence systems produce inventions: An alternative model for patent law at the 3A era. Cardozo Law Review, 39, 2215-226

IX. APPENDIX A. CORPUS CODING MATRIX

The appendix summarizes the coded corpus used for the descriptive analysis. It links the literature, policy materials, selected legal decisions, and sector guidance documents to the coding categories used in the study.

Appendix Table A1. Coded corpus matrix used for descriptive analysis.

ID	Source group	Jurisdictional focus	Primary issue	Governance code
S1	Scholarly study	International	Code-based authorship	Authorship
S2	Scholarly study	United States	Non-human creativity	Authorship
S3	Scholarly study	United States	Machine creativity ownership	Ownership
S4	Scholarly study	Comparative	Computer-generated works	Originality
S5	Scholarly study	International	Robot-generated art	Human authorship
S6	Scholarly study	European Union	Machine-learning authorship	Originality
S7	Scholarly study	United States	Authors and machines	Human contribution
S8	Scholarly study	United States	Artificial creativity	Doctrinal gap
S9	Scholarly study	United States	Copyright dilemma	Copyrightability
S10	Scholarly study	European Union	Machine-learning outputs	Ownership
S11	Scholarly study	International	Creative adversarial networks	Digital-art creativity
S12	Scholarly study	International	Computer art	Cultural interpretation
S13	Scholarly study	International	Autonomy and intention	Artistic meaning
S14	Scholarly study	International	Art and creativity	Legal interpretation
S15	Scholarly study	International	Creative industries	Human-centered use
S16	Scholarly study	United States	Fair use and learning	Training data
S17	Scholarly study	United States	Text and data mining	Training data
S18	Scholarly study	European Union	TDM exceptions	Training-data transparency
S19	Scholarly study	United States	Fair learning	Licensing balance
S20	Scholarly study	European Union	Data and creative production	IP limits
S21	Scholarly study	United Kingdom/EU	Producer and consumer roles	Input-output link
S22	Scholarly study	European Union	Protected outputs	Human contribution

ID	Source group	Jurisdictional focus	Primary issue	Governance code
S23	Scholarly study	European Union	Regulatory balance	Policy design
S24	Scholarly study	International	IP governance	Public interest
S25	Scholarly study	United States	Creative machines	Moral/economic interests
S26	Scholarly study	United States	Originality and automation	Originality
S27	Scholarly study	United States	Ownership allocation	Policy coherence
S28	Scholarly study	United Kingdom	Digital copyright disruption	Creative labor
S29	Scholarly study	United States	Platform-mediated creativity	Platform responsibility
S30	Scholarly study	European Union	Authorship doctrine	Copyrightability
S31	Scholarly study	United States	Market incentives	Reform criteria
P1	Institutional policy material	United States	Copyrightability guidance	Human authorship
P2	Institutional policy material	United States	Training-data legal issues	Training-data legitimacy
P3	Institutional policy material	European Union	Training-content summary template	Transparency
P4	Institutional policy material	United Kingdom	Copyright and automated systems consultation	Licensing policy
P5	Institutional policy material	International	Copyright and frontier technologies discussion	Policy coordination
C1	Registration or court material	United States	Zarya of the Dawn	Human selection/arrangement
C2	Registration or court material	United States	Théâtre D'opéra Spatial	Generated material disclaimer
C3	Registration or court material	United States	Thaler v. Perlmutter	Human authorship boundary
G1	Platform/sector guidance	Creative-industry sector	Output-use terms	Contractual clarity
G2	Platform/sector guidance	Creative-industry sector	Creator disclosure and attribution	Transparency
G3	Platform/sector guidance	Creative-industry sector	Complaint and rights-reservation procedures	Creator remedies