

Framework for Responsible Use of **AI in Art** **Authentication**

Developed by Art Recognition, a leader in the AI art authentication industry and a commercial entity based in Switzerland, and co-authored with the Center for Art Law, a nonprofit organization based in the United States.



ART RECOGNITION

Carina Popovici
Noah Charney



Center for art law

Irina Tarsis
Atreya Mathur

Executive Summary

These Guidelines (the “Guidelines”) offer a framework for the responsible, ethical, and transparent use of artificial intelligence (AI) in the authentication of artworks. They aim to provide art market participants, scholars, cultural institutions, collectors, and technology developers with shared principles and practical steps to ensure that AI complements, rather than disrupts, established authentication practices. By articulating core values such as transparency, accountability, and collaboration, the Guidelines seek to safeguard artistic legacies, scholarly integrity, and cultural heritage.^[1]

Responsible use of AI in art authentication demands transparency from developers and active collaboration with art experts.

As artificial intelligence (AI) continues to transform industries from healthcare to finance, its potential to assist in the authentication of artworks is increasingly recognized and utilized. In the art world, where questions of authorship, authenticity, moral rights, and attribution can shape reputations, scholarship, market value, and museum collecting practices, the stakes are uniquely high. Historically, authentication has rested on three interdependent pillars: scientific analysis, connoisseurship, and provenance research. Together, these approaches provide a foundation of reasonable assurance in determining an object’s attribution.^[2]

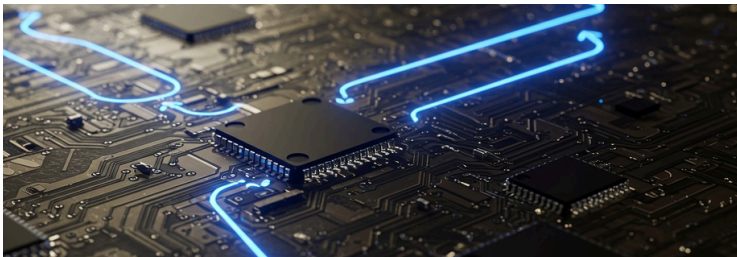


^[1] AI tools are inherently probabilistic and must be understood as complementary to, not a substitute for, connoisseurship, provenance research, and scientific analysis. All AI outputs or research should be critically assessed within the broader context of expert review, scholarly interpretation, and applicable legal and ethical frameworks. Terms with * are explained in the Glossary in the Appendix.

^[2] Sarah Cascone, “Expert Forgery-Spotter James Martin to Head Sotheby’s Scientific Research Department,” Artnet (Dec. 5, 2016) available at <https://news.artnet.com/market/james-martin-sothebys-scientific-research-771905>

In the AI-Art sector, AI technologies currently offer significant new capabilities. They can detect stylistic patterns and anomalies invisible to the human eye, process and compare vast datasets, and provide probabilistic insights to support attribution decisions. Yet, while the computational power of AI represents a significant advance in technology, its application must be approached with a combination of enthusiasm, critical scrutiny, and ethical caution. Responsible use of AI in art authentication demands transparency from developers and active collaboration with art experts.

Developers and service providers must disclose the methodologies, model architectures, and data sources underpinning their AI systems. This information should be communicated in clear and accessible language so that non-technical experts can understand and critically evaluate AI outputs. Training datasets must be curated with scholarly oversight, drawn from verified and lawful sources, respect intellectual property rights, and undergo regular review and updates.



Beyond technical considerations, the integration of AI raises significant legal and ethical questions. Responsible AI development requires embedding ethics and compliance at every stage of the AI lifecycle, from data curation and model design to interpretation, disclosure, and public communication. This includes avoiding overstatement^[3] of AI findings, respecting intellectual and cultural property rights, and keeping up with regulatory change.

Central to these Guidelines is the emphasis on collaboration between AI developers and art experts, starting from the earliest stages of AI design and training. Mutual respect and open exchange are essential to responsibly integrate AI into established authentication practices.

Finally, these Guidelines are intended to be dynamic and adaptive, evolving alongside technological innovation, legal developments, and academic research. Through sustained review and a commitment to transparency, accountability, and collaboration, AI can enrich authentication processes while honoring artistic reputations, scholarly rigor, and cultural values.

^[3] AI findings include the results of the AI image analysis test, which is comprised of a “heat map” showing how much of an artist’s hand is likely present in different sections of the painting analyzed, the percentage likelihood that the artist in question is behind the work, and other elements provided by the AI image analysis firm to the client.

1. Scope and Applicability

The integration of AI into art authentication demands a reexamination of traditional workflows, ethical boundaries, and professional responsibilities. These Guidelines provide a comprehensive framework for the responsible creation, deployment, and interpretation of AI systems within the field of art authentication. Their purpose is not only to suggest best practices, but to help build a durable culture of transparency, accountability, and collaboration among all stakeholders who engage with these emerging technologies.



1.1 Objectives

The Guidelines are intended to serve two interconnected objectives. First, they seek to ensure that the design and development of AI models for art authentication prioritize scientific validity, ethical responsibility, and active collaboration with art historians and art historical research. The creation of such models must be undertaken with an awareness of their limitations in matters of connoisseurship and historical context, as well as with an understanding of the interpretive nuances that differ significantly from more conventional AI application in other fields. Human-AI collaboration is at the core of AI use in art authentication.

Second, they aim to encourage responsible and informed adoption of AI technologies by art professionals, including connoisseurs, provenance researchers, forensic analysts, curators, lawyers, and market participants. AI findings should be seen as a valuable additional layer of analysis that enhances traditional methodologies, not as a replacement for them. These Guidelines aim to help ensure that AI is deployed in ways that honor both scholarly rigor and responsible stewardship of objects and associated data.

The foundational principles of transparency, scientific rigor, ethical responsibility, and human-AI collaboration must guide both the development of robust AI authentication models and the responsible, well-informed integration of AI into art authentication practices.

They are intended to apply across the entire lifecycle of an AI authentication system. This lifecycle begins with conceptualization and design. It proceeds to model training and validation, to deployment and interpretation of results, communication of findings to diverse stakeholders, and ultimately, to the practical application of these results.

1.2 Audience and Scope

The Guidelines are created for use across a variety of sectors. The intended audiences for these Guidelines include, but are not limited to two main groups.

AI Developers. These are individuals and organizations designing and deploying AI models specifically intended for use in art authentication. This includes commercial AI firms, academic research groups, and independent data scientists.

Art Experts. These include individuals and institutions identified as connoisseurs, forensic specialists, provenance researchers, museums, collectors, auction houses, galleries, insurers, attorneys, as well as others within the art sphere who may use AI authenticity analyses as part of the authentication process.

The Guidelines are intended to be global in scope. While regulatory environments may vary from country to country, the fundamental principles articulated here are universally applicable. Art, after all, transcends national boundaries, and so too must the standards governing the new technologies that interact with it.

1.3 Collaboration with Art Experts

AI does not replace human judgment — it extends it, through collaborative insight grounded in respect for expertise.

AI technologies offer powerful new tools for analyzing artworks, but their true value in authentication arises only when they are integrated into established expert practices. The future of authentication lies not in replacing human judgment with algorithms, but in fostering a symbiosis where technological precision and human expertise complement and enhance one another.

Collaboration between AI Developers and Art Experts depends on mutual respect. Developers must avoid overstating the capabilities of their systems and recognize the essential role of art historical knowledge from human experts in both creating datasets and interpreting outputs. Likewise, Art Experts should approach AI findings with informed openness, viewing them as additional and complementary empirical insights rather than threats to their authority. A culture of reciprocal attention and shared inquiry allows both audiences to contribute meaningfully to the authentication process.

An Integrative Model: AI and traditional methods should be used in tandem with traditional methods—including expert analysis (connoisseurship), technical imaging, material studies, and provenance research—bring contextual depth, historical sensitivity, and interpretive nuance that AI cannot replicate. Conversely, AI can process massive amounts of visual information with speed and consistency unattainable by human beings, potentially spotting micro-level patterns invisible to even the trained eye. A balanced approach means interpreting AI findings alongside expert assessments, not accepting them uncritically. Authenticity conclusions should emerge through dialogue between machine outputs and human reasoning, resulting in more transparent and reasoned decisions.

Structured Partnerships and Capacity Building: Effective integration requires early and structured collaboration between developers, connoisseurs, authentication committees, provenance researchers, and conservation scientists. Expert input should guide the selection of training data, the identification of stylistic markers, and the design of tools tailored to real-world challenges. Authentication committees of the sort run by artist foundations, or groups of scholars preparing updated editions of *Catalogues Raisonnés*, can incorporate AI findings, such as heatmaps or confidence scores, as advisory materials, similar to technical imaging reports. Capacity building is also essential. Developers should provide accessible training, user manuals, and workshops to help art professionals understand what AI can and cannot do, how to interpret its outputs, and how to integrate them responsibly into decision making. Certification programs could help standardize best practices and foster a knowledgeable user base. Conversely, developers must learn from art historians to appreciate how restoration, overpainting, and workshop practices can shape authentication outcomes.



2. Technical Standards

2.1 Transparency

Transparency is a core ethical requirement. Unlike many commercial applications of AI, where end users may not need, or want, to understand how a model works, the world of art authentication operates in a setting where credibility, scholarly scrutiny, and legal and market implications demand a higher standard of disclosure. The technology must not function as a “black box,” in the sense of the general public being unaware of how something works, and therefore being suspicious of what they do not understand. While the competitive nature of commercial AI ventures may require firms to protect proprietary information, this should not be used as an excuse for complete opacity.

The foundational principles of transparency, scientific rigor, ethical responsibility, and human-AI collaboration must guide both the development of robust AI authentication models and the responsible, well-informed integration of AI into art authentication practices.

Instead, AI developers must aim for methodological openness, providing sufficient insight into their tools so that users can make informed, critical use of the results. Ethical AI development demands a balanced approach: retaining unique intellectual property while still revealing enough about the methodology to permit meaningful scrutiny. Think of this as analogous to peer-reviewed publication in the sciences: while a pharmaceutical company may not reveal its entire manufacturing process, it is still expected to publish trial data, methods, and outcome statistics.

Peer review is another critical element of transparency. Developers should seek external validation of their AI models through joint publications with academic partners, or independent replication studies. AI systems should not be released or relied upon without vetting by qualified third parties. The goal is credibility. An AI model must be demonstrably grounded in sound scientific practice.

2.2 Documentation

At a minimum, AI firms or developers working in art authentication should maintain detailed documentation of their models. This documentation should outline the model's intended function, the type of machine learning architecture used (e.g., convolutional neural networks, transformers, ensemble methods), the rationale for selecting that architecture, and a history of model development, including versioning and updates. It should also describe the input data format, the type of output generated, and the performance metrics used to interpret results.

AI developers should also produce plain-language summaries to explain the system's operation to non-technical stakeholders. Ideally, firms should make these summaries publicly available on their websites or through white papers, and update them regularly as models evolve.

2.3 Communication of Limitations

The limitations of the technology must also be transparently communicated and documented. Developers must clearly communicate what their AI technologies cannot do. For example, a system trained only on high-resolution images may not perform reliably when analyzing photographs of damaged works. Models trained only on a subset of an artist's oeuvre—say, late-period oils but not early drawings, should not be used to make inferences outside that scope. If AI is, for example, unable to analyze “drip paintings” due to the random nature of how they are created (as opposed to a trained and repeatable artistic hand), then this should be clearly communicated and duly reflected in AI analysis of such works.

2.4 Explainability

Explainability is the ability to make the operations and conclusions of an AI model understandable to human users. It is a central pillar of AI deployment in art authentication. In this highly specialized and often contentious field, it is not sufficient for an AI system to simply issue a prediction or classification. Stakeholders must be able to understand how and why an AI model reached its conclusion. Explainability allows for critical assessment of the validity of an AI-made decision, weighing it against other forms of evidence, and determining its appropriate role in decision-making processes. Only then can its insights be used constructively.

Without clear and comprehensible explanations, users risk treating AI outputs as oracular pronouncements—accepting results uncritically or rejecting them out of mistrust. Neither extreme serves the cause of responsible authentication. Properly explained AI findings invite critical engagement, allowing art professionals to question, corroborate, or challenge their own and machine's conclusions within the broader framework of evidence and expert judgment.

Moreover, explainability fosters accountability. If an AI system's conclusions can be explained and scrutinized, errors can be identified and biases can be corrected. In contrast, if conclusions are opaque, mistakes are likely to go unnoticed and uncorrected, leading to potential injustices in attribution decisions, legal disputes, or financial transactions.

In the long term, developing highly explainable AI tools will also contribute to building trust in the broader use of AI across the art world. When AI systems are seen not as inscrutable authorities but as collaborative partners—transparent in their methods, clear in their reasoning, and humble in their claims—their integration into art historical and forensic workflows will be both more effective and more ethically sound.

Best practices for explainability include:

Visual Explanations: visual explanations, such as *heat maps*, show which regions of an artwork most influenced the AI's decision. For example, a heat map might reveal that the model paid particular attention to the rendering of a particular anatomical detail. These visual cues enable art experts to compare the AI's focus with their own observations and expectations, fostering a dialogue between machine analysis and human connoisseurship.

Probabilistic Outputs with Context: AI systems should provide probabilistic results rather than issuing binary judgments like “authentic” or “not authentic.” These should be accompanied by explanations of what they represent and how they relate to the training dataset's scope and composition.

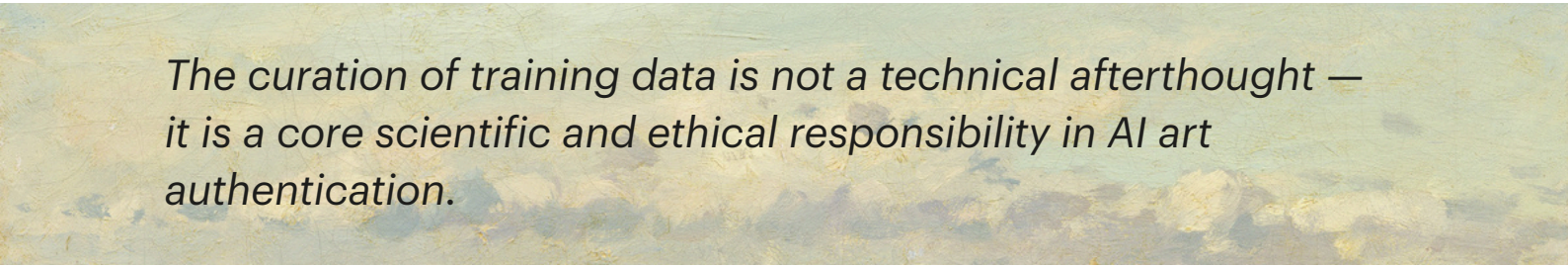
Model Behavior Transparency: developers should have model behavior transparency. They should disclose how the model behaves under different conditions. For example, how it responds to variations in image resolution, lighting conditions, or the type of camera used. Understanding the model's strengths and vulnerabilities helps users apply it appropriately.

3. AI Models and Creation

3.1 Training Datasets

Important factors in training datasets used for AI art authentication are quality, relevance, and integrity. Machine learning systems are only as good as the data they are fed. If training data is incomplete, biased, or flawed, the AI's conclusions will be correspondingly unreliable.

Properly training datasets requires positive examples, or in other words works that are accurately attributed to a given artist, for the AI to draw upon. Wherever possible, these examples should be cross-referenced with established Catalogue Raisonnés, which are widely recognized as the scholarly gold standard for cataloging an artist's authentic oeuvre. This ensures that the AI is learning from consensus attributions and minimizes the risk of "teaching" the model based on misattributed or dubious pieces. In some cases, even Catalogue Raisonné entries are disputed among scholars. When uncertainty exists, those works should either be excluded from the training set or flagged appropriately during model development. This requires the hand-in-glove teamwork of human experts and the AI. The art expert must consider works with their knowledge and experience and be there to help the AI, as well as using the AI as a tool to help their analysis and interpretation. The goal is to avoid introducing ambiguity at the foundational level of the AI's learning process. Therefore, strict quality control is essential.



The curation of training data is not a technical afterthought — it is a core scientific and ethical responsibility in AI art authentication.

Equally important is to include negative examples, or works that are not authentic creations of the artist in question. Ideally, negative training datasets should include non-authentic works such as high-quality imitations by contemporaries or followers; workshop productions not executed by the artist's hand; later copies and forgeries produced with intent to deceive; or works that were once attributed to the artist but have since been deattributed.

Both positive and negative examples must be confirmed and relevant. As human-AI collaboration is at the forefront of AI use in art authentication, these examples should be assembled with the guidance of art historians, curators, conservation scientists, art experts and other professionals.

As always, transparency is key. Firms and researchers should disclose information such as:

- the total number of positive and negative examples used;
- art historical literature included;
- material restrictions;
- gaps or biases in the dataset; and
- art experts consulted in the selection of the data and works used in the training of these datasets.

Datasets must also be kept current. As new scholarship emerges, (for example, the discovery of previously unknown works, changes in attribution status, or updates to Catalogue Raisonné entries), datasets must be revised accordingly. Thus there should be regular review and updates of AI systems.

Investing in the creation of robust, well-documented, and sound datasets, AI developers lay the essential groundwork for systems that can contribute meaningfully, respectfully, and credibly to the profoundly human endeavor of art authentication.



3.2 AI Models and Architecture

The architecture of an AI model is the blueprint that defines how it processes information, learns patterns, and generates predictions. This plays a crucial role in its performance, reliability, and interpretability. In art authentication, model architecture must be chosen with careful consideration of the field's particular challenges such as small dataset sizes, high variability between works, nuanced stylistic and material signals, and the need for interpretability by experts. Artworks are the product of an individual stylistic evolution, historical context, and, often, intentional deception by imitators and forgers. Recognizing these subtleties requires an architecture capable of analyzing such intricate and domain-specific features.

AI models for art authentication must be purpose-built and designed to meet the field's unique complexities.

It is critical that AI models deployed for art authentication are purpose-built, meaning they are specifically designed for the task of evaluating an artwork's authenticity. AI architecture originally built for other domains is unlikely to perform reliably when applied directly to the subtleties of fine art.

Furthermore, multimodal architectures that combine visual analysis with textual data—drawing on art historical sources that reference an artist’s style, influences, recurring subjects and motifs, as well as historical context—could be explored. Whatever the choice, developers should avoid a one-size-fits-all mentality and instead pursue context-sensitive solutions tailored to the specific challenges of art authentication.



Explainability is not just a technical feature—it is an ethical imperative for building trust and accountability in AI art authentication.

Rigorous testing practices are extremely important in assessing whether an AI model can appropriately be used in the art authentication process. Developers should use cross-validation to minimize bias and improve model robustness. In the case of tenfold cross validation, for example, the available dataset is divided into ten subsets: the model is trained on nine subsets and tested on the remaining one, repeating the process ten times so that each subset serves as a test set once.

Model performance should be assessed using specific metrics. Three potential metrics to be used are accuracy, precision and recall.

Accuracy: Accuracy measures the proportion of correct predictions out of total cases.

Precision: Precision assesses the accuracy of positive predictions, or how many artworks the model identified as authentic are actually authentic. This metric is particularly important where there is a risk of a false positive. For example, mistakenly validating a forgery.

Recall: Recall measures the model's ability to correctly identify all authentic works, or how many of the genuine artworks the model successfully recognized. Recall becomes critical in scenarios where false negatives (failing to detect a genuine piece) could lead to the loss or misclassification of culturally or financially significant artworks. Developers should define specific terms and metrics such as what constitutes a “positive” or “negative” prediction in their system, whether confidence thresholds are applied, and how these thresholds are determined.

Model evolution must be part of an ongoing commitment. As machine learning research advances, new architectures, training techniques, and optimization strategies regularly emerge. AI developers working in the art domain should monitor these innovations closely and be prepared to retrain, fine-tune, or even redesign their systems to take advantage of proven improvements.

4. Legal and Ethical Considerations

The use of AI in art authentication raises complex legal and ethical questions that extend well beyond technical accuracy. Errors, misuses, or misinterpretation of AI outputs can affect the reputations of artists and scholars, influence the art market, and impact the perception and preservation of cultural heritage. The goal should be to cultivate a broader culture of ethical awareness among AI developers, art professionals, and end users.

This includes fostering open dialogue about the ethical implications of AI findings and providing training and continuing education on the responsible use of AI in art trade. By embedding ethics into every aspect of AI authentication, from dataset curation and model design to interpretation and public communication, developers and users can help ensure that technology enriches, rather than diminishes, careful cultural stewardship.

Responsible AI in art authentication requires legal care, ethical sensitivity, and a commitment to cultural stewardship.

4.1 Intellectual Property, Copyright, and Licensing

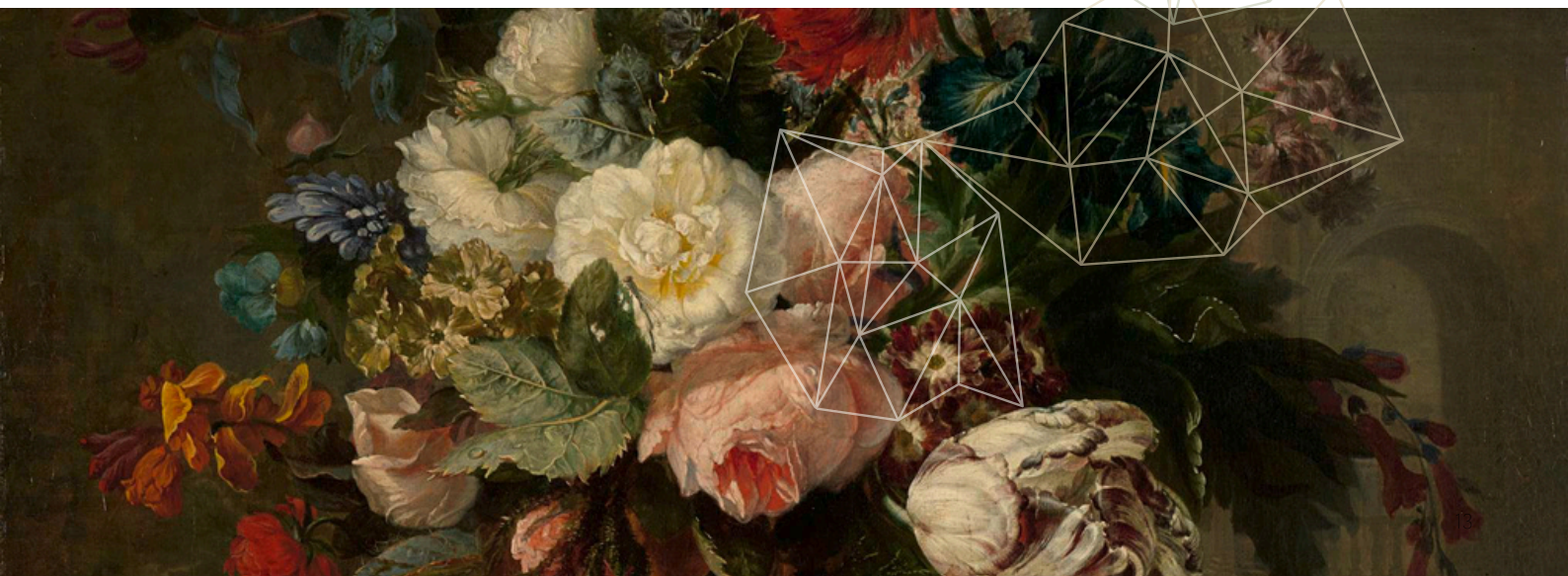
Central to these considerations are copyright and intellectual property (IP) issues. AI systems rely on training datasets composed of images, photographs, scans of artworks or written scholarship, many of which are protected under copyright, neighboring rights, or moral rights. Using images of signed, attributed, or otherwise identifiable works without permission can constitute infringement. Likewise, using scholarship by art historians and related researchers may infringe on their rights. For instance, training an AI system on high-resolution photographs of contemporary paintings from a private collection without explicit authorization could expose both the developer and the institution using the tool to legal liability. Similarly, reproducing AI-generated reports or visualizations derived from copyrighted works without proper licensing may violate the rights of the artist or estate.

Ethical AI deployment therefore requires securing relevant releases and permissions, licensing proprietary content including images, and providing proper attribution. Both developers and users must respect the IP of creators while mitigating potential legal risks. For example, any public sharing or publication of AI-generated analyses should be carefully reviewed to ensure that the copyright of the underlying training content is not infringed in relevant jurisdictions.

4.2 Responsibilities of Developers and Users

Both AI developers and end users, including museums, auction houses, and private collectors, carry critical responsibilities to ensure the ethical and lawful use of AI in art authentication. Developers must obtain all images and training data through lawful means, securing appropriate licenses or permissions from artists, estates, and institutions. Usage terms, whether for research-only purposes or commercial applications, should be clearly defined and communicated. Users of AI reports must respect associated copyright and licensing restrictions when using AI outputs. Even if an AI report suggests a possible attribution, any public sharing of the report or its visual analyses may implicate the copyright of the underlying images used to train the AI, and should only occur under proper circumstances (e.g. permissible use, licensed release, court order, etc).

Equally important is the shared responsibility to prevent misuse of AI tools. AI outputs should not be presented as definitive evidence of authenticity. Even high-probability results require corroboration through traditional methods, such as expert assessments, provenance research, and scientific analyses like pigment testing or light spectrum imaging. Developers should implement safeguards, including internal policies, disclaimers, user agreements, and contractual limitations on how outputs may be marketed or cited. These practices ensure that AI complements human expertise rather than replacing it and preserves the credibility and integrity of authentication processes.



4.3 Cultural Heritage and Ethical Obligations

Beyond copyright, AI developers and users must consider cultural property and heritage laws. Artworks from Indigenous, colonized, or historically contested contexts demand special stewardship. Training datasets containing such works should be compiled in consultation with cultural representatives. Using images of sacred or culturally sensitive objects without consent, even if publicly accessible online, could violate both ethical norms and international heritage conventions concerning traditional cultural expressions. Developers must embed these ethical considerations into dataset curation, model design, and output interpretation.

4.4 Privacy and Cybersecurity

Authentication datasets often include sensitive provenance documents, condition reports, or unpublished images containing personal data about collectors, artists, or institutions. Misuse or breaches of this information could violate privacy laws such as the European Union's GDPR, the California Consumer Privacy Act or other relevant national or state laws. Developers must implement encryption, controlled access, and anonymization where possible to protect confidentiality. These measures safeguard stakeholders' privacy and preserve trust in AI-assisted authentication processes.

AI model developers must be aware of and comply with data retention regulations. To this end, they should implement policies that either justify the retention of certain data or establish clear data purging timeframes.

4.5 AI Outputs as Evidence

AI-generated analyses are increasingly considered in legal disputes, insurance claims, and provenance challenges, but they are not definitive evidence on their own. Courts typically require transparency about the AI system's methodology, training data, and interpretability. For example, if there is a disputed attribution of a certain work of art, AI analyses should be accompanied by disclosures of how the model was trained, the sources of its data, and known error rates. Overreliance on AI without corroborating human expertise can undermine the credibility of authentication.



4.6 Compliance and Accountability

AI in art authentication operates within a rapidly evolving legal and regulatory environment. Developers and institutions must stay informed of changes in copyright and intellectual property law, licensing frameworks, national and state regulations, privacy rules, and AI-specific legislation such as the [EU AI Act](#). (See eg. [Article 13](#) and [Article 26](#)) They must also consider evidentiary standards for expert testimony, which may affect the admissibility of AI outputs in legal or scholarly contexts.

To ensure legal and ethical conformity, organizations should establish clear accountability mechanisms and governance structures. These mechanisms define who is responsible for monitoring compliance, reviewing training datasets, validating model methodologies, and overseeing the responsible use of AI outputs. A key role within this structure would potentially be a compliance officer: a professional, often supported by legal counsel, tasked with ensuring that AI activities adhere to applicable laws, ethical guidelines, and internal policies. Compliance officers would oversee risk management, implement policies to prevent misuse or infringement, and coordinate with developers and users to maintain transparency and integrity throughout the AI lifecycle.

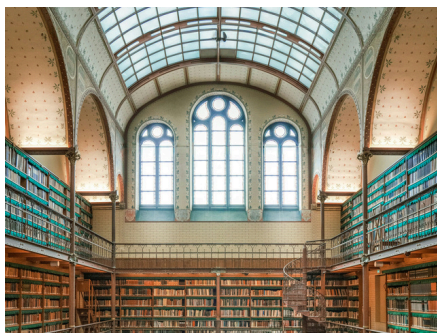
Compliance is not a one-time obligation; it must be proactive, continuous, and integrated into all stages of AI development and deployment. By combining rigorous oversight, clear accountability, and transparent processes, institutions can ensure that AI tools enhance authentication responsibly, support human expertise, and respect the legal and moral rights of artists, collectors, and cultural stakeholders.



5. Review and Updates

The rapid evolution of both AI technologies and the broader fields of art historical research, cultural heritage law, and market practice demands that best practice guidelines for AI in art authentication be dynamic, adaptive, and responsive. Therefore, a key commitment of these guidelines is regular, structured review and updating, informed by the collective experiences and evolving needs of all stakeholders in the authentication ecosystem.

To remain credible and effective, AI authentication guidelines must evolve alongside the disciplines they support and the professionals and stakeholders they serve.



5.1 Scheduled Review

At a minimum, these Guidelines should be formally reviewed every two to three years. Such scheduled reviews must be proactive, anticipating technological, methodological, and regulatory changes rather than reacting only to crises or disputes. Reviews should consider:

Advances in machine learning architectures and AI capabilities relevant to visual analysis, anomaly detection, or pattern recognition.

Emerging authentication methodologies, including novel imaging technologies, pigment analysis, and provenance network mapping.

Developments in legal and regulatory frameworks, including copyright law, intellectual property law, data privacy, and AI-specific regulations such as the EU AI Act.

Feedback and practical experiences from a diverse range of stakeholders, including art historians, conservators, museums, auction houses, collectors, legal experts, and technical specialists.

5.2 Ongoing Feedback and Field Engagement

In addition to scheduled reviews, continuous feedback mechanisms are critical. These enable the Guidelines to remain responsive to the evolving needs of the field and to the practical realities of AI implementation. Effective feedback mechanisms may include:

Comment period that invites feedback from users and developers for improvements and clarification of the Guidelines.

Annual forums or symposia focused on AI in art authentication, showcasing case studies, technical developments, and ethical reflections.

Online portals for practitioners to submit concerns, suggest improvements, or propose new areas of consideration for future revisions.

Advisory councils or ethics committees comprising technologists, art historians, conservators, legal scholars, and market participants tasked with monitoring field developments and recommending updates.

Public consultation periods allow review and commentary from the wider community, including underrepresented regions, smaller institutions, and independent scholars.



5.3 Principles for Guideline Updates

All updates should adhere to the following principles:

- **Transparency:** Proposed changes and the rationale behind them should be communicated publicly and made open for discussion.
- **Accountability:** Authors and contributors responsible for drafting updates should be clearly identified, and mechanisms must exist for stakeholders to challenge or question revisions.
- **Continuity and Coherence:** While adaptability is essential, updates should preserve the core principles of transparency, collaboration, ethical responsibility, and respect for human expertise.
- **Documentation:** Each version should be archived, dated, and accompanied by a change log detailing modifications, additions, or removals of sections and the reasoning behind the changes.



These Guidelines are intended not as a static rulebook but as an best practices framework, capable of evolving alongside technological advances, legal shifts, and the professional practices of authentication. By embracing this iterative process, AI in art authentication can maintain credibility, foster trust among stakeholders, and ensure that human expertise remains central to decision-making.

Regular updates, combined with proactive engagement and inclusive consultation, will enable the field to respond effectively to emerging challenges and opportunities while safeguarding the integrity of cultural heritage.

6. Conclusion

Artificial Intelligence offers significant potential for the field of art authentication. By leveraging sophisticated pattern recognition and high-speed data analysis, AI systems can contribute powerful new tools to a process historically governed by scientific analysis and human expertise. Yet with great potential comes equally great responsibility. The future of AI in art authentication will not be determined solely by technological innovation, but also by the ethical, thoughtful, and collaborative choices made today.

These Guidelines propose a path forward that emphasizes transparency, accountability, collaboration, and respect for the irreplaceable value of human judgment. AI should not be viewed as a final arbiter of authenticity but rather as an important and powerful aid, that can extend and enrich human inquiry while honoring the interpretive complexity and historical sensitivity that define the field of art history.

At its best, AI can help uncover hidden insights, identify subtle patterns, and bring empirical rigor to aspects of authentication that have long been shrouded in subjectivity. At its worst, however, AI can introduce new risks: the erosion of scholarly standards, the temptation to replace nuanced judgment with mechanized certainty, the commercialization of dubious claims of authenticity, and the alienation of those without technical training. These dangers are not hypothetical, they are real and must be confronted proactively through ethical frameworks like the one outlined here.

The successful integration of AI into art authentication requires a culture of critical engagement, not technological determinism. Stakeholders must be empowered to ask hard questions about how AI systems are trained, what assumptions they encode, what data they rely on, and how their outputs are interpreted. Developers must commit to ongoing transparency, robust validation, and continual education.

By following these Guidelines and committing to a spirit of openness, collaboration, and responsible innovation, the art world can harness the benefits of AI while safeguarding the values that have long made art a central and irreplaceable part of human civilization. The future of art authentication can be one in which technology and tradition work in concert to protect, understand, and celebrate the treasures of our shared cultural past.



Appendix: Glossary of Technical Terms

Model

The final trained AI system that can evaluate new data (in this case, artworks) based on what it has learned during training.

Model Architecture

Model architecture is the overall design of an AI system. It defines the family the model belongs to (like CNN or Transformer) and the internal design choices, such as how many layers it has and how those layers connect.

Training Dataset

It's the full pool of data (images, labels, and notes) that teaches the model patterns. It is usually divided into subsets for training and testing, yet the entire pool is often referred to as the training dataset.

Heatmaps

Generally, heatmaps use color to show intensity across an image or grid. In the context of AI, heatmaps are used to show which parts of the image have a higher contribution to the decision-making process of the model.

Confidence Score

Confidence score is a calibrated prediction: a Number between 0 and 100% that aims to reflect real-world likelihood for authenticity. Higher scores suggest a greater likelihood, but they're probabilistic estimates, not certainties.

Technical Imaging Reports

Technical imaging reports are conservation-style studies (X-ray, infrared reflectography, ultraviolet fluorescence, and pigment or binder analysis). They are used to reveal a work's materials and underdrawing for authenticity assessment.

Convolutional Neural Networks

CNN is an image-focused AI architecture that learn pattern-matching filters that slide over an image. It's loosely inspired by how our visual system responds to small regions of sight.

Transformers

Transformers are a flexible AI architecture that learns which parts of the input to focus on. First built for language, they now power everything from text to images, and they're the breakthrough behind LLMs like ChatGPT.

Ensemble Methods

Ensemble methods combine several models (like a panel of experts) into one prediction. The group is usually more accurate and stable than any single model.

Multimodal Architectures

These architectures fuse multiple data types (image, text, sometimes audio) into a single understanding. By combining information from different domains, they capture relationships that a single-modality model would miss.



ART RECOGNITION

Art Recognition is a Swiss ArtTech company created in 2019 with the mission to make art authentication transparent, objective, and accessible. Using an in-house developed AI technology, the company empowers art collectors, museums, galleries, auction houses, art dealers and experts to authenticate artworks and make informed decisions with scientific confidence. By bridging Artificial Intelligence with art history and enhancing the scope of traditional connoisseurship, Art Recognition is redefining authenticity for the digital age.

www.art-recognition.com



Center for art law

Center for Art Law is a Brooklyn-based research and education nonprofit that offers resources and programming for the advancement of a vibrant arts and law community. Launched in 2009 as an online resource, the Center evolved into a go-to leader in the field that provides learning opportunities to a worldwide community. Today, the Center is the only independent art law entity in the United States dedicated to writing, gathering, and sharing law and visual arts information for the benefit of artists, students, lawyers, academics, and many more.

www.itsartlaw.org