

Creator Attribution Amendment

Fixing Article 50(2) of the EU AI Act for Human Authorship

Core Proposal Summary

1. Executive Summary

- **The Missing Link:** Existing technology and regulations identify the software provider but fail to offer a standardized option to assert human authorship in assistive workflows.
- **The Legislative Proposal:** A targeted policy obligation requiring AI tool providers to expose a user-controlled, voluntary authorship assertion mechanism.
- **Core Feasibility:** Leverages existing authenticated user sessions and open provenance standards (C2PA/CAWG) without requiring new technology or infrastructure.

2. The Current Problem: Provider-Centric Provenance

- **Technical Reality vs. User Need:** Current provenance manifests identify the tool/issuer (e.g., OpenAI, Google) but leave the human user completely absent.
- **Corporate Risk-Avoidance:** Fear of regulatory fines under Article 50 forces providers into blanket, binary tagging for all outputs regardless of edit scale.
- **Failure to Implement Assistive Exceptions:** Why the statutory exemption for assistive editing in Article 50(2) remains functionally unused in real-world software.

3. The Structural Gap: The Missing User Option

- **The Unconnected Building Blocks:** Authenticated session IDs exist, C2PA manifests exist, and legal exemptions exist - yet no standard interface option connects them.
- **Absence of User Agency:** Users are deprived of a standardized setting to bind their active account identity to their creative direction.
- **The Legal Bottleneck:** Why the gap is purely an unoffered software feature, not a technical or technological limitation.

4. The Policy Proposal: Obligation to Expose an Option

- **Scope of the Mandate:** Obliging AI providers to offer a voluntary, user-controlled toggle (e.g., *"Attach Account Identity to Content Credentials"*).
- **Legal Target:** Targeted amendment to Regulation (EU) 2024/1689 (EU AI Act), Article 50(2) implementation frameworks.
- **Strictly Voluntary & Privacy-Preserving:** Preserves complete user autonomy and GDPR compliance by maintaining unsigned defaults.

5. Technical Implementation Mechanics (Using Existing Assets)

- **Session-Bound Identity:** Leveraging active OAuth/account sessions as the implicit notary layer.
- **C2PA / CAWG Integration:** Populating standard schema fields (c2pa.edited, Author Assertions) using the provider's existing signing certificate.
- **Workflow Execution:** Co-signing the manifest to output a dual-verified record upon user export.

6. Expected Systemic Benefits

- **For Creators:** Seamless, one-click ability to stand behind their work and claim human authorship.
- **For AI Providers:** Full regulatory compliance for Article 50(2) without incurring complex edge-case liabilities.
- **For Regulators & Downstream Platforms:** Traceable accountability, improved content signals, and a practical realization of the law's original intent.

Executive Summary

The Missing Link

Current AI provenance implementation creates a critical failure point in digital publishing workflows: while cryptographic metadata and watermarking systems successfully identify the AI platform or software provider, they completely fail to record or verify the human creator directing the process. In human-led, assistive editing workflows, the human author remains entirely absent from the provenance record.

The Legislative & Technical Gap

Article 50(2) of Regulation (EU) 2024/1689 (EU AI Act) explicitly establishes a statutory exemption for AI systems performing an "assistive function for standard editing". However, to avoid regulatory fines, AI providers default to blanket, binary tagging on all processed media. This occurs not because of technological limitations, but because platforms do not provide users with a standardized interface setting to bind their active account identity to their creative output.

The building blocks: authenticated user sessions, open provenance standards (C2PA/CAWG), and legal exemptions - already exist. The gap is strictly the absence of a required user option to connect them.

The Proposed Solution

The **Creator Attribution Amendment** introduces a targeted policy obligation requiring AI tool providers to expose an optional, user-controlled interface setting upon export. By toggling this setting, an authenticated user can voluntarily co-sign the output's provenance manifest using their active session credentials. This binds the platform's existing signing certificate to a clear assertion of human authorship and assistive editing (c2pa.edited).

Core Ecosystem Impact

- **For Creators:** Restores agency and protects European moral rights (*droit moral*) by enabling artists and professionals to voluntarily claim primary human direction over their work.
- **For AI Providers:** Delivers a clear, standardized framework to fulfill the assistive editing intent of Article 50(2) without incurring complex liability edge cases.
- **For Regulators & Downstream Systems:** Establishes verifiable, identity-backed attribution signals that differentiate transparent human creation from untraceable, automated media.

The Current Problem: Provider-Centric Provenance

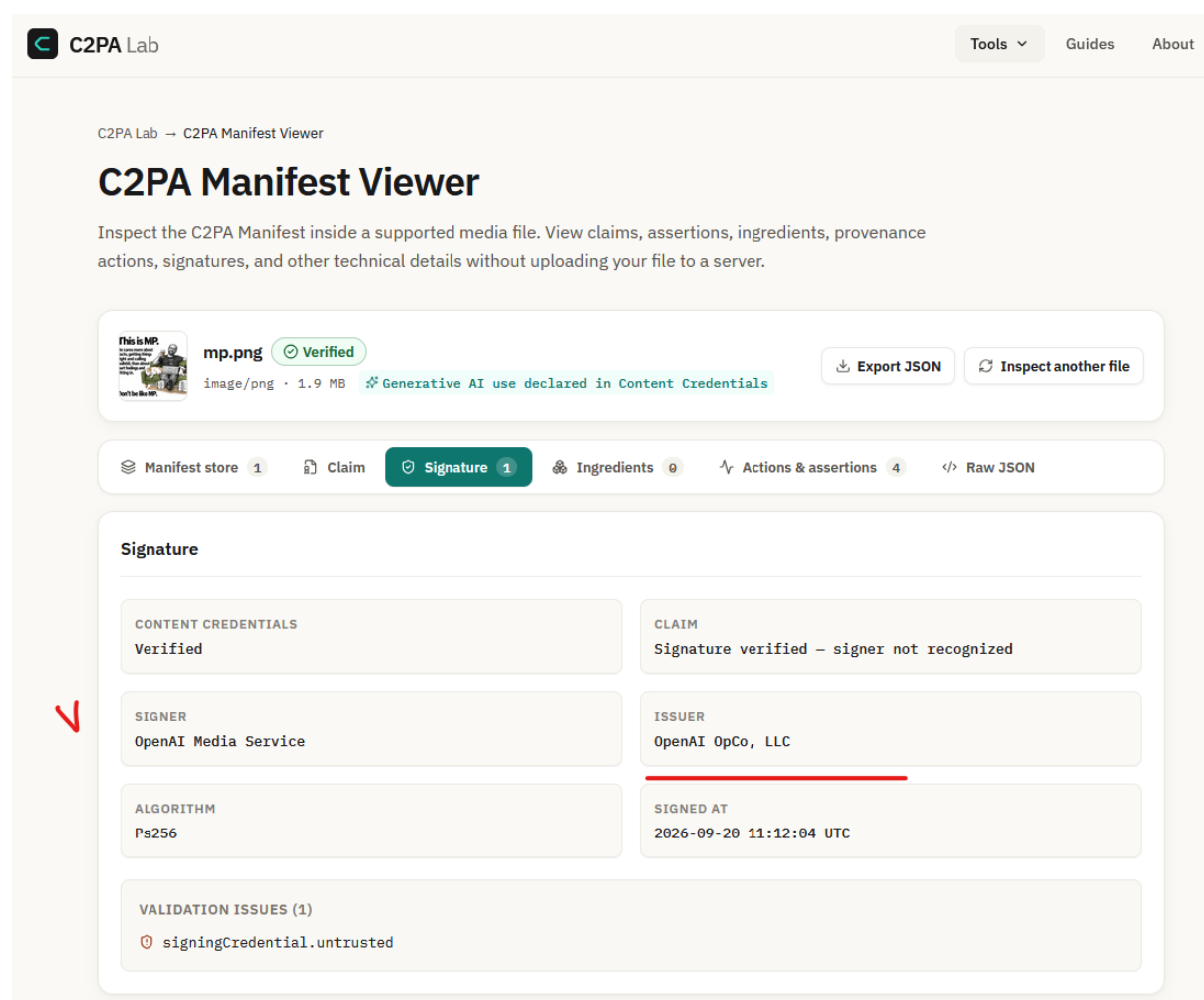


Figure 1: The Provenance Gap.

Current C2PA implementation in professional AI tools. The manifest records the platform (OpenAI) as the sole signer and issuer, while the human creator is entirely absent. The `signingCredential.untrusted` error demonstrates how current "blanket" compliance efforts create validation failures that lead to automated down-ranking and content suppression.

Technical Reality vs. User Need

Current provenance implementations prioritize platform identity over human authorship. When an author uses an AI tool to perform assistive edits—such as focus repair, lighting adjustments, or localized canvas extensions—the resulting file’s manifest records only the AI platform (e.g., OpenAI, Google, etc.) as the primary signer and issuer. The human creator who supplied the original work, directed the software, and made the creative decisions is completely absent from the cryptographic signature and metadata claims.

Corporate Risk-Avoidance and Blanket Tagging

Under Regulation (EU) 2024/1689 (EU AI Act), AI providers face substantial financial penalties for failing to detect and mark artificially generated or manipulated media. Because evaluating whether a specific edit constitutes "substantial alteration" versus minor "assistive editing" introduces potential compliance ambiguity, software companies adopt a zero-risk compliance strategy. If a generative algorithm processes any pixel array, the system applies an indiscriminate, blanket watermark or signal. This blunt enforcement transforms what should be a nuanced attribution log into a binary flag.

The Structural Failure of Statutory Exemptions

Article 50(2) of the EU AI Act was designed with an explicit statutory carve-out to protect human creation: transparency obligations do not apply where AI performs an assistive function for standard editing. In practice, this exemption is functionally unusable. Because distribution gatekeepers (search engines, content marketplaces, and social networks) execute automated scans reading only raw pixel watermarks or generic platform signatures, they cannot detect workflow context. As a result, human-directed, assistively edited works are systematically classified as fully synthetic media, subjecting creators to automated reach restrictions, algorithmic suppression, and the loss of their moral right to attribution (*droit moral*).

The Structural Gap: The Missing User Option

The Unconnected Building Blocks

The failure to protect human authorship under Article 50(2) is not caused by a lack of technology. The fundamental infrastructure already exists across three distinct layers:

- **Identity Layer:** AI platforms already maintain authenticated user sessions through account credentials, single sign-on (SSO), and active user subscriptions.
- **Provenance Layer:** Open standards such as C2PA (Coalition for Content Provenance and Authenticity) and CAWG (Creator Assertions Working Group) provide established schema fields to log specific actions (c2pa.edited) and attach creator assertions.
- **Legal Layer:** Regulation (EU) 2024/1689 (EU AI Act) Article 50(2) provides the legal principle exempting assistive editing from blanket synthetic media marking.

The structural gap exists because no regulatory standard obliges software providers to connect these three layers into a functional user feature.

Absence of User Agency in Export Workflows

In current software architectures, users are denied agency over the provenance metadata embedded in their exports. When an author saves or exports a project that incorporates assistive AI processing,

the platform software automatically writes its own organizational key as the sole signing entity. The user interface provides no mechanism, checkbox, or configuration setting that allows the creator to assert their creative direction or attach their authenticated session identity to the export.

A Policy Gap, Not a Technological Limitation

Because big tech providers design compliance mechanisms strictly to minimize corporate regulatory liability, they default to blanket platform-level tagging. Without a legal mandate requiring platforms to expose a user-controlled attribution option, providers have no incentive to build user-facing signing controls. The barrier preventing human creators from asserting their rights under EU law is entirely an unoffered interface setting.

The Policy Proposal: Obligation to Expose an Option

Scope of the Regulatory Mandate

To bridge the structural gap, this proposal establishes a targeted policy obligation for providers of generative AI systems and editing software operating within the Union. Providers must expose a voluntary, user-controlled interface mechanism that allows authenticated users to assert human authorship and workflow context during content generation or export.

This mandate does not require platforms to alter algorithmic distribution, rank content, or adopt proprietary architectures. Instead, it requires providers to make existing session-authentication and cryptographic signing capabilities directly accessible to the user as an explicit preference setting.

[Default User Preference UI]

[✓] Attach Account Identity to Content Credentials

"Binds your authenticated account ID to assistive edits as primary author to preserve attribution and legal exemptions under EU law."

Targeted Amendment to EU AI Act Implementation

This policy operates as a binding enforcement rule under Regulation (EU) 2024/1689 (EU AI Act), specifically clarifying the operational requirements for Article 50(2):

1. **Mandatory Interface Provision:** Providers subject to Article 50 obligations must integrate a clear, accessible toggle within their software interface enabling users to opt into identity-bound attribution.
2. **Standardized Contextual Schema:** When enabled, the software must write the active user session identity alongside the specific action type (c2pa.edited or equivalent) directly into the provenance claim, rather than issuing a generic, platform-only synthetic tag.
3. **Harmonized Compliance Recognition:** Provenance manifests carrying platform-authenticated user assertions for assistive operations must be recognized by Market Surveillance Authorities and downstream systems as fulfilling the Article 50(2) assistive editing exemption.

Voluntary Design and Privacy Compliance

The proposed mechanism strictly respects user autonomy and fundamental privacy rights:

- **Voluntary Opt-In:** The assertion feature is non-mandatory for the end user. Users retain full freedom to export media without attaching identity metadata.
- **GDPR Harmonization:** By leveraging existing account relationships and giving users explicit granular control over display credentials (e.g., pseudonymous account ID vs. verified public name), the mechanism complies fully with Regulation (EU) 2016/679 (GDPR) data minimization principles.
- **No Unsolicited Tracking:** Provenance binding occurs locally or at the point of user-initiated export, preventing unsolicited tracking while giving creators a verifiable seal when they choose to stand behind their work.

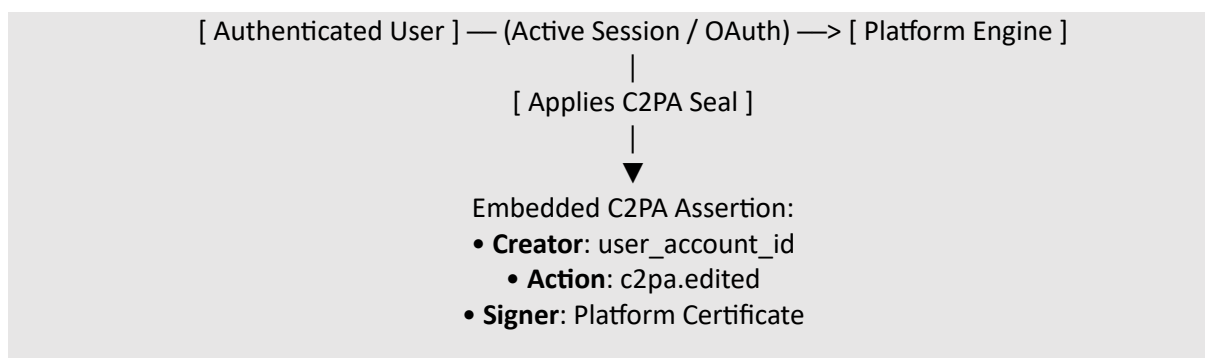
Technical Implementation Mechanics (Using Existing Assets)

Session-Bound Identity as the Primary Notary

The technical implementation requires no new cryptographic infrastructure or user-managed private keys. AI software platforms already require users to maintain active, authenticated sessions via Single Sign-On (SSO), OAuth, or paid account credentials.

Under this model, the AI provider's server acts as a trusted notary:

1. **Active Authentication:** The user logs in and initiates an editing or generation workflow.
2. **Preference Check:** The platform checks the user's interface preference setting (e.g., *"Include Account Identity in Content Credentials"*).
3. **Server-Side Signing:** Upon export, the provider applies its standard, trusted cryptographic signing certificate to seal the file's provenance manifest.



Integration with C2PA and CAWG Schema Standards

The co-signed manifest utilizes existing open standards to record workflow context without breaking compatibility:

- **Action Labeling:** Instead of applying an unnuanced root generation tag (c2pa.created), the platform writes an assistive editing action claim (c2pa.edited or compositeWithTrainedAlgorithmicMedia).
- **Creator Assertions:** Utilizing the Creator Assertions Working Group (CAWG) specification within the C2PA framework, the platform populates standard creator fields with the user's account identifier or chosen display name.
- **Dual-Signed Record:** The output file carries a dual-verified manifest. The platform's signature guarantees the technical integrity of the file history, while the embedded user assertion establishes the human author's creative direction.

Downstream Automated Verification Logic

By providing a standardized, dual-signed manifest, downstream distribution gatekeepers (search engines, social networks, and content platforms) can evaluate provenance programmatically:

1. **Detection:** Automated scanners detect the presence of generative AI code or low-level pixel signals.
2. **Manifest Inspection:** The scanner reads the accompanying C2PA container.
3. **Context Evaluation:** Rather than defaulting to a generic "100% Synthetic AI" label, the verification filter recognizes a platform-verified Creator assertion attached to an assistive operation (c2pa.edited).
4. **Resolution:** The system applies the Article 50(2) exemption, preserving the asset's organic reach, platform indexing, and author attribution.

Systemic Benefits & Multi-Stakeholder Impact

For Human Authors and Content Creators

- **Protection of *Droit Moral*:** Upholds the legal right to claim primary authorship and creative direction over modified assets under European copyright frameworks.
- **Mitigation of Algorithmic Suppression:** Helps preventing search engines, social media gatekeepers, and distribution networks from automatically down-ranking assistively edited work through unnuanced, binary "AI Media" labels.
- **Zero Technical Friction:** Bypasses the need for individual creators to manage personal cryptographic private keys or digital certificates by leveraging existing authenticated platform sessions.

For Generative AI System Providers

- **Clear Compliance Framework:** Provides a standardized, legally defensible mechanism to fulfill the "assistive editing" exception established in Article 50(2) of the EU AI Act.
- **Low-Cost Implementation:** Eliminates the need to develop proprietary evaluation models to determine whether an edit is "substantially altering" or "assistive"—delegating the attribution assertion directly to the authenticated user.
- **Interoperable Ecosystem Alignment:** Leverages existing open standards (C2PA/CAWG) and server-side signing pipelines without introducing new system overhead.

For Distribution Platforms and Regulators

- **High-Quality E-E-A-T Signals:** Downstream networks gain access to verified, account-bound metadata that separates reputable, human-directed creative work from automated AI spam.
- **Traceable Accountability:** Eliminates "anonymous AI cover" by binding actions to authenticated user accounts, ensuring creators accept responsibility when publishing content.
- **Realization of Legislative Intent:** Fulfills the dual purpose of the EU AI Act by providing robust transparency against deceptive synthetic media while preserving human creative rights across the Union.

Implementation Scope & Future Research

Primary Technical Readiness (Visual Media)

The operational mechanics outlined in this proposal draw directly upon established, production-ready C2PA and CAWG specifications. These standards are currently most mature for visual media formats, including static images, digital graphics, and video files, where persistent binary container headers are already widely supported.

Extensibility Across Modalities

Article 50(2) of Regulation (EU) 2024/1689 applies universally to all forms of AI-assisted outputs, including text, synthetic voice, and audio. While file-based media can seamlessly embed metadata containers, text-based generation workflows present distinct technical challenges (such as plain-text copy-pasting) that strip away traditional binary metadata headers.

Directive for Regulatory Guidance

To ensure comprehensive enforcement across all media types, regulatory bodies and standards organizations should initiate targeted technical research and stakeholder consultations. This will help establish standardized, persistent user-assertion protocols for non-file text workflows, ensuring human authorship rights are uniformly protected as digital provenance standards evolve.