C2PA and eIDAS 2.0 - a Reference Framework for Content Verification and Protection in the EU

Stefan Wagenpfeil[1], Joyce Amy Kolenda, Waldemar Nauvomets, Tim Ternovik and Jakob Wenzig

Department of Computer Science, Software Engineering and AI
PFH Private University of Applied Sciences, Weender Landstrasse 3-7, 37073 Göttingen, Germany

Abstract

This paper presents an integrated approach for trustworthy multimedia evidence in legal contexts, combining C2PA content provenance, eIDAS 2.0 trust services, and the MMIR reference framework. We demonstrate how cryptographically verifiable provenance, qualified signatures, and privacy-preserving identity credentials can be linked to images, audio, and video, enabling robust chains of custody and legal admissibility. The prototype shows that this standards-based solution delivers machine-verifiable authenticity, cross-border legal recognition, and operational resilience, setting a new benchmark for digital evidence in judicial and forensic applications.

Keywords

Multimedia Information Retrieval, Digital Signatures, Authenticity, Validation

1. Introduction

The increasing prevalence of digital multimedia evidence in legal proceedings, ranging from criminal investigations to civil litigation, has fundamentally transformed the evidentiary landscape. While images, audio, and video recordings have long been regarded as powerful forms of proof, the advent of advanced generative AI and sophisticated editing tools has rendered traditional notions of authenticity and integrity precarious [1, 2, 3]. Deepfakes, synthetic media, and undetectable manipulations challenge courts, law enforcement, and legal practitioners to distinguish genuine evidence from fabricated or altered content [1].

In response to these challenges, two complementary frameworks have emerged at the intersection of technology and law: the Coalition for Content Provenance and Authenticity (C2PA) [4] and the revised European eIDAS Regulation (eIDAS 2.0) [5]. C2PA provides a technical standard for embedding cryptographically verifiable provenance metadata, so-called content credentials, directly into multimedia assets. This enables tamper-evident documentation of the entire lifecycle of a media object, from capture through editing to dissemination [4]. In parallel, eIDAS 2.0 establishes a comprehensive legal and technical infrastructure for digital identity, qualified electronic signatures, seals, timestamps, and attribute attestations across the European Union, culminating in the introduction of the European Digital Identity Wallet (EUDI Wallet) [6].

The convergence of C2PA and eIDAS 2.0 offers a unique opportunity to construct end-to-end trustworthy chains of custody for digital evidence. By binding cryptographically secure provenance data to legally recognized digital identities and qualified trust services, it becomes possible to elevate the evidentiary value of multimedia assets to a level suitable for judicial scrutiny [7]. This integration not only addresses technical requirements for integrity and authenticity but also fulfills the stringent legal standards for admissibility and probative force, particularly in European courts [5, 7].

TRI-IVIS - AVI 2026
*Corresponding and main author.
s.wagenpfeil@pfh.de (S. Wagenpfeil)
https://www.pfh.de (S. Wagenpfeil)
0000-0003-2100-7589 (S. Wagenpfeil)

This article situates the discussion within a broader research program on trustworthy multimedia and AI integration [7], substantially contributing to the use of trustworthy and trusted media in Multimedia Information Retrieval (MMIR) [7] as conceptual foundations. We focus specifically on the requirements, architectures, and practical implications of combining C2PA and eIDAS 2.0 for forensic and legal media processing. Through this lens, we aim to delineate the technical, legal, and organizational prerequisites for establishing robust, transparent, and court-admissible multimedia evidence workflows in the era of synthetic media. Following the Nunamaker methodology [8], this article consists of four major parts: first, we will show the state of the art in science and technology, as well as related work in section 2. Second, a detailed modeling of the extension is given in section 3, which is then demonstrated by a proof-of-concept implementation in section 4. An evaluation of the results achieved is presented in section 5.

2. Legal Frameworks and Requirements, State of the Art

The reliability and admissibility of digital multimedia evidence in legal and forensic contexts depend on a robust interplay between technical standards and regulatory frameworks. In the European Union, three pillars are particularly relevant: the eIDAS 2.0 Regulation, which governs digital identity and trust services (see subsection 2.1); the C2PA standard, which provides technical mechanisms for content provenance and is presented in subsection 2.2; and in subsection 2.3 the EU Artificial Intelligence Act, which introduces new obligations for transparency and trustworthiness in AI-generated content. Together, these frameworks establish the legal and technical conditions necessary for trustworthy multimedia processing in judicial and investigative settings.

2.1. eIDAS 2.0: Digital Identity and Trust Services

The revised eIDAS Regulation (Regulation (EU) 2024/1183) [5] is the cornerstone of digital trust in the European Union. It introduces the European Digital Identity Wallet (EUDI Wallet), a secure and standardized digital wallet that allows citizens and organizations to store and present digital credentials, such as qualified electronic signatures, seals, timestamps, and attribute attestations [6]. These credentials are legally recognized across all EU member states and are designed to ensure that digital transactions—including the submission and authentication of multimedia evidence—are as trustworthy as their physical analogues.

For example, a forensic investigator documenting a crime scene can use a mobile device to capture images and sign them with a qualified electronic signature issued via their EUDI Wallet. This signature, recognized under eIDAS as legally equivalent to a handwritten signature, provides a presumption of authenticity and integrity in court. If the investigator's official role is relevant, an attribute attestation—such as proof of law enforcement status—can be attached to the evidence, further strengthening its legal standing. Qualified trust service providers (QTSPs) are responsible for issuing and managing these credentials under strict supervision, ensuring that signatures, seals, and timestamps are generated according to rigorous security standards. In practice, a timestamp applied to a digital photograph by a QTSP is presumed accurate and reliable, unless proven otherwise—a crucial factor in legal disputes where the timing and authorship of evidence are contested.

The harmonized framework of eIDAS 2.0 also facilitates cross-border recognition of digital evidence. For instance, a video recording authenticated and timestamped in Germany using eIDAS-compliant services must be accepted with the same legal force in France or Italy. This interoperability is particularly valuable in transnational investigations or litigation, where evidence often crosses jurisdictional boundaries.

2.2. C2PA: Technical Provenence and Content Credentials

While eIDAS 2.0 provides the legal provenance for digital trust, the Coalition for Content Provenance and Authenticity (C2PA) [4] addresses the technical challenge of establishing and verifying the provenance of

multimedia assets. The C2PA standard enables the embedding of cryptographically verifiable provenance metadata—referred to as content credentials—directly within digital files such as images, audio, and video.

A C2PA manifest records the entire lifecycle of a media asset, including details about the capture device, the time and location of acquisition, and a tamper-evident log of all subsequent edits or transformations. For example, a police body camera equipped with C2PA-compliant firmware can automatically generate a signed manifest at the moment of recording. Each time the footage is accessed, edited, or transferred, the manifest is updated with a new cryptographic signature, creating a transparent and immutable chain of custody.

This technical provenance is invaluable in legal contexts. Consider a scenario where a video of a traffic accident is submitted as evidence. If the video's C2PA manifest shows an unbroken chain of signatures from the moment of capture to its presentation in court, and each signature is linked to a verified identity (potentially via eIDAS credentials), the court can be confident in both the integrity of the file and the legitimacy of each actor in its history. Conversely, any attempt to tamper with the file or its metadata would break the cryptographic chain, immediately alerting forensic analysts and legal professionals to potential manipulation.

C2PA's interoperability is further enhanced by its adoption among major hardware and software vendors. Modern smartphones, professional cameras, and editing suites are increasingly supporting C2PA, making it feasible to implement provenance tracking from the point of capture through to final dissemination. This widespread support is essential for ensuring that provenance information is preserved even as files move across different platforms and jurisdictions.

2.3. The EU AI Act: Trustworthiness and Transparency in AI Systems

The regulatory landscape for trustworthy multimedia is further shaped by the EU Artificial Intelligence Act (AI Act), which introduces new obligations for transparency, accountability, and risk management in AI systems. The AI Act requires that providers of AI-generated content—including images, audio, and video—implement mechanisms for clear disclosure and traceability, especially when such content is used in contexts where authenticity and provenance are critical [9, 7].

For instance, under the AI Act, platforms distributing synthetic media must ensure that such content is clearly labeled as AI-generated. This requirement complements the technical capabilities of C2PA, which can embed metadata indicating whether a file was created or modified by an AI system. In legal and forensic settings, this means that courts and investigators can reliably distinguish between authentic and synthetic evidence, reducing the risk of manipulation or deception.

The AI Act also mandates risk assessments and documentation for high-risk AI systems, including those used in law enforcement, judicial proceedings, and critical infrastructure. Providers must maintain detailed records of system design, training data, and operational history, which can be linked to C2PA manifests and eIDAS credentials to create a comprehensive audit trail. This integrated approach not only enhances the trustworthiness of AI systems but also ensures compliance with legal standards for evidence admissibility and probative value.

In summary, the combined effect of eIDAS 2.0, C2PA, and the EU AI Act is to establish a multi-layered framework for trustworthy multimedia processing. Legal, technical, and organizational safeguards work in concert to ensure that digital evidence is authentic, reliable, and admissible in judicial and investigative contexts.

2.4. Related Work

The trustworthy integration of multimedia and artificial intelligence has become a central research topic in computer science, legal informatics, and digital forensics. Over the past decade, the rapid evolution of generative AI, multimodal information retrieval, and provenance technologies has fundamentally altered the landscape of digital evidence, content authenticity, and user trust. This section provides an

overview of the current state of the art, highlighting seminal works and recent advances that underpin the technical and legal frameworks discussed in previous sections.

Early research in multimedia information retrieval (MMIR) established the foundations for indexing, searching, and managing large collections of images, audio, and video [10, 11]. These systems evolved from keyword-based approaches to content-based and multimodal retrieval, leveraging advances in computer vision, audio analysis, and natural language processing. The introduction of deep learning and large-scale datasets such as ImageNet [12] enabled significant improvements in feature extraction, semantic understanding, and cross-modal querying. All such MMIR systems follow a similar processing workflow and component structure, which is shown in Figure 1.

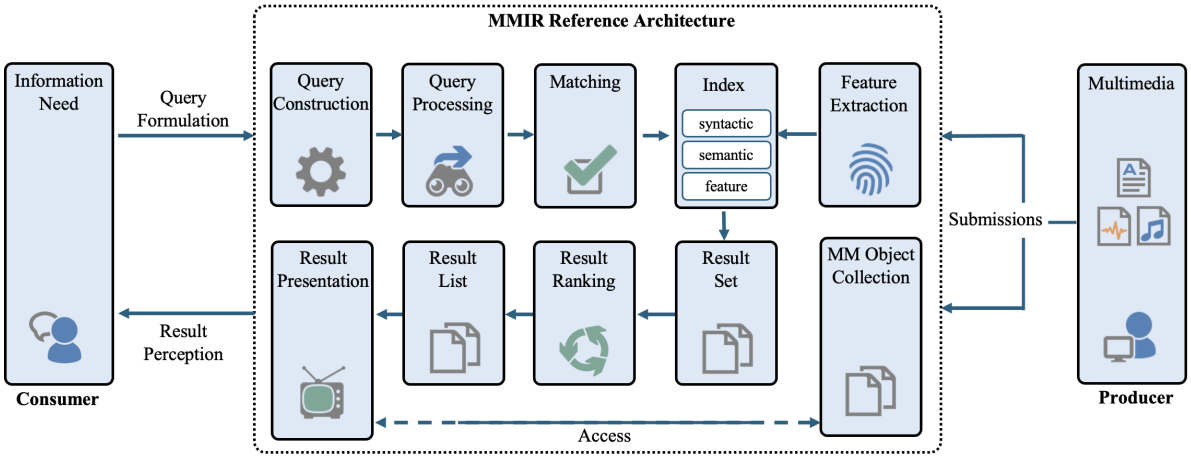

Figure 1: Multimedia Information Retrieval reference architecture. Own visualization based on [10].

Recent surveys have documented the transition from unimodal to multimodal retrieval-augmented generation (MRAG), where AI models combine stored knowledge with generative capabilities across text, image, audio, and video modalities [13, 14]. These approaches have demonstrated state-of-the-art performance in tasks such as semantic search, evidence attribution, and cross-modal recommendation, but have also exposed new challenges related to modality bias, robustness, and provenance.

The proliferation of synthetic media and deepfakes has raised urgent questions about authenticity, trustworthiness, and governance [1]. Technical solutions such as digital watermarking, cryptographic signatures, and provenance tracking have emerged as safeguards against manipulation and misinformation. The C2PA standard [4] represents a major advance in this area, providing interoperable mechanisms for embedding tamper-evident content credentials in multimedia assets. Complementary efforts, such as Google's SynthID watermarking for AI-generated text and video, further enhance the ability to detect and label synthetic content.

On the regulatory front, the EU AI Act and eIDAS 2.0 have introduced new requirements for transparency, risk management, and legal admissibility of digital evidence [5]. These frameworks mandate machine-readable marking of synthetic media, qualified electronic signatures, and robust chains of custody, aligning technical innovations with legal standards for evidence and accountability. In recent work, a *Generic Multimedia Analysis Framework (GMAF)* including data structures for the representation of *Multimedia Feature Graphs (MMFG)* has been developed as a reference architecture [15, 16, 17] supporting effective, efficient, and explainable MMIR processes.

2.5. Industry Support

In this section, a brief overview of the current support in various industries is given. Both hard- and software-related aspects are described. In subsection 2.5.1, a selection of the most prominent camera vendors and hardware is given, which lay the foundation of C2PA industry support. Subsection 2.5.2 outlines software solutions both for C2PA and eIDAS 2.0, and subsection 2.5.3 show examples for platforms exposing content credentials.

2.5.1. Camera Vendors and Hardware

The increasing maturity of C2PA as a provenance standard is also reflected in its growing adoption by industry, particularly among leading manufacturers of imaging and media-production technology. This industrial uptake serves as a practical complement to the legal and technical frameworks discussed in Sections 2.1–2.4, underscoring the feasibility of embedding trustworthy provenance directly into professional production workflows. Among the early and most influential adopters is Sony, which has played an active role in shaping the C2PA ecosystem as a member of the initiative's Steering Committee [4].

In 2024, e.g., **Sony** released the first C2PA-compliant firmware updates for still-image provenance, extending Content Credentials to widely used mirrorless cameras such as the Alpha 1 (v2.00), Alpha 7S III (v3.00), and Alpha 7 IV (v3.00) [18]. These updates represented a significant step toward integrating tamper-evident metadata at the moment of capture, enabling photographers to bind cryptographically verifiable provenance information—including capture context, device information, and edit history—to original image files in a manner fully consistent with the C2PA v2.2 specification. This expansion of provenance support aligns with the broader trend noted in Section 2.2, where hardware-level integration of C2PA manifests is essential for establishing continuous chains of custody. A major technological milestone followed in 2025 with Sony's release of the PXW-Z300, the first commercially available C2PA-enabled video camera [19]. By supporting video-level Content Credentials—including the embedding of 3D depth metadata, temporal signatures, and structured provenance assertions—the PXW-Z300 established a new benchmark for video authentication. This advancement is especially relevant for journalism and investigative reporting, where courts and news organizations increasingly require transparent, verifiable media provenance. Through integration with the CI Media Cloud, Sony's collaborative workflow platform, journalists can inspect embedded provenance data and incorporate it directly into editorial verification pipelines, thereby operationalizing the provenance concepts introduced by C2PA. In addition to new hardware, Sony has systematically expanded C2PA-aligned video authentication across its broader product portfolio. Since mid-2025, the Alpha 1 II, Alpha 9 III, and the Cinema Line models FX3 and FX30 have supported video-level Content Credentials, bringing a substantial portion of Sony's professional and semi-professional ecosystem into alignment with emerging authenticity requirements [20]. Further firmware releases planned for the second quarter of 2025 extend these capabilities to the Alpha 7R V, Alpha 7 IV, and Alpha 1, reinforcing the trend toward portfolio-wide provenance integration. Continuing this trajectory, Sony announced that the Alpha 7S III, which already provides C2PA-compliant still-image credentials, will receive full video authentication support in 2026 [18]. Taken together, Sony's roadmap demonstrates that C2PA is transitioning from a conceptual framework to a widely supported industrial standard. The continuous expansion of both still-image and video provenance functionality across camera families reflects the increasing demand for authenticated media in legal, journalistic, and forensic contexts. This aligns closely with the regulatory landscape outlined in the preceding sections: as eIDAS 2.0 and the AI Act establish stronger legal obligations for trustworthiness, industry adoption of C2PA provides the technical substrate needed to implement these obligations at scale. Consequently, the convergence of regulatory frameworks and industrial implementation marks a critical turning point toward mainstream, end-to-end trustworthy media production.

In addition to Sony's early integration efforts, also **Leica** has positioned itself as one of the most influential pioneers in bringing C2PA-compliant provenance technology into high-end photographic workflows. Leica was the first manufacturer worldwide to ship a professional full-frame camera with natively embedded Content Credentials, namely the Leica M11-P, which marked a decisive turning point in the commercial adoption of cryptographically verifiable provenance in the photography sector [21]. Unlike earlier experimental or software-based prototypes, the M11-P introduced hardware-anchored provenance support at the moment of capture, fully aligned with the data-model principles laid out by the C2PA specification [4, 22]. Through this integration, Leica became the first manufacturer to

Figure 2: C2PA Content Credential settings in the menu of a Sony α7 IV.

integrate CAI's Content Credentials seal directly into the camera's firmware, enabling tamper-evident claims for creators, journalists, and digital archivists. Following this foundational step, Leica invested in a broader strategic rollout of provenance-enabled hardware across multiple product lines. By 2025, the company expanded C2PA-based Content Credentials to additional models including the Leica M11-D, Leica SL3-L, Leica M EV1, and the Leica Q3 Monochrom, thereby establishing one of the most comprehensive product portfolios with native provenance support in the photographic industry [23]. Leica's approach parallels broader developments discussed in Sections 2.2 and 2.3: the rapid rise of synthetic media, the regulatory pressures introduced by the AI Act for transparency labeling, and the need for trusted metadata infrastructures [1, 7]. Through its rapid expansion of supported devices, the company directly contributes to creating a resilient provenance ecosystem across multiple market segments, from analog-styled manual-control cameras to hybrid photography-videography platforms. A distinguishing factor of Leica's implementation is the company's emphasis on ecosystem integration. Unlike many manufacturers that rely exclusively on third-party viewers or platform-level validation features, Leica provides a tightly coupled environment where users can inspect provenance information either via the official Content Credentials viewer operated by the Content Authenticity Initiative or directly through the Leica FOTOS App [21]. This mobile application not only displays the embedded manifest and associated signature details but also uses an intuitive presentation layer that mirrors the disclosure guidelines established by the C2PA UX framework. As a result, photographers and journalists can confirm authenticity and review provenance chains directly during fieldwork, long before files enter newsroom workflows, DAM systems, or evidentiary chains. From a compliance and governance perspective, Leica's proactive support reflects the broader movement toward provenance-aware imaging pipelines and anticipates regulatory requirements that may emerge across the EU under eIDAS 2.0 and the AI Act. While Leica's current devices primarily focus on still-image provenance, their architecture is technically aligned with the C2PA JUMBF-based manifest container model, which positions the company to expand toward multi-modal provenance—potentially including short-form video, HDR imaging stacks, or computational photography pipelines. By prioritizing native, camera-level provenance over cloud-only solutions, Leica effectively reduces attack surfaces such as metadata stripping or re-encoding, two vulnerabilities frequently highlighted in C2PA threat mappings [4]. Taken together, Leica's trajectory demonstrates the growing momentum of C2PA as not merely a

technical standard but a foundational layer for trustworthy, creator-centric imaging workflows. As more manufacturers align with provenance-first design principles, Leica's early leadership underscores how high-end camera systems can play a central role in securing the authenticity of visual media in journalistic, forensic, archival, and cultural-heritage contexts.

Alongside Sony and Leica, **Nikon** was among the earliest camera manufacturers to publicly commit to provenance-based authenticity frameworks. As early as 2021, Nikon joined both the Content Authenticity Initiative (CAI) and the Coalition for Content Provenance and Authenticity (C2PA), becoming one of the first major imaging companies simultaneously represented in both organizations [24]. This dual membership positioned Nikon as an early advocate for standardized, tamper-evident metadata across professional photography workflows. In October 2024, Nikon announced a firmware update for the Nikon Z6 III, introducing support for C2PA-based content verification, and signalling the company's ambition to expand its capabilities in the field of media authenticity [25]. However, in September 2025, Nikon temporarily suspended C2PA certification for the Z6 III due to a technical issue identified during validation. Certification remained paused until the underlying problem could be resolved, illustrating the engineering challenges inherent in integrating cryptographically secured provenance at device level [26]. Nikon's workflow also differs notably from those of Sony and Leica. To enable C2PA functionality, users must first register with the Nikon Image Cloud and link their camera to the service. In contrast to systems where device-embedded credentials are issued directly by the manufacturer, Nikon requires photographers to obtain an external C2PA-compliant certificate from a third-party trust provider, which must then be uploaded to the Nikon Image Cloud to activate Content Credentials [27]. Nikon itself does not issue the necessary signing certificates, resulting in a more modular—but also more setup-intensive—provenance pipeline.

Beyond the companies already discussed, several additional camera manufacturers have begun orienting themselves toward provenance-aware imaging pipelines, even if they have not yet implemented full C2PA integration. **Canon** has participated in authenticity-related industry dialogues and previously explored in-camera verification approaches through proprietary metadata integrity systems on EOS-series hardware. While Canon has not released C2PA-compliant firmware, its R-series mirrorless architecture and metadata models remain technically compatible with future JUMBF-based provenance extensions [28]. **Fujifilm** has likewise invested in strengthening in-camera transparency and metadata fidelity, reflecting its strategic focus on computational imaging across the X-Series and GFX-Series platforms. Although no commercial Fujifilm camera yet embeds C2PA manifests, the company has aligned itself with broader authenticity initiatives and continues to develop robust EXIF pathways and RAW-processing disclosure frameworks that could evolve into provenance-compatible pipelines [29]. **Panasonic**, driven by its strong presence in professional filmmaking, particularly through the Lumix GH and Lumix S product lines, has emphasized codec transparency, metadata persistence, timecode robustness, and workflow accountability. While Panasonic has not formally introduced C2PA integrations, it has repeatedly signaled support for emerging authenticity standards as part of its professional-video ecosystem development [30]. Across these manufacturers, a broader movement is visible: although direct C2PA support is still limited, their technical roadmaps increasingly align with the requirements for standardized, tamper-evident content provenance mechanisms.

2.5.2. Software

As trustworthy media workflows increasingly require interoperability between camera-embedded provenance and post-production environments, software vendors play a decisive role in maintaining Content Credentials throughout the editing chain. **Adobe**, as both a founding member of the Content Authenticity Initiative (CAI) and a leading contributor to the C2PA standard [4], has become a central actor in ensuring that provenance is preserved, extended, and verifiable across creative workflows. This is particularly critical in light of modern AI-based editing tools, which can substantially alter visual content by adjusting lighting, modifying scenes, or even replacing photographed objects

entirely. To prevent such edits from obscuring or destroying provenance, Adobe integrates Content Credentials directly into its industry-standard applications Photoshop and Lightroom, embedding C2PA-compliant watermarks, manifests, and metadata into image files in an immutable manner [31]. Through the Adobe Authenticity (Beta) ecosystem, Adobe guarantees compliance with C2PA standards by maintaining verifiable provenance records throughout the entire editing process [32]. Beyond basic provenance preservation, the beta platform enables users to enrich Content Credentials with additional attribution-relevant information—such as verified creator identities, professional account links, or social-media identifiers—increasing transparency and reducing the risk of unauthorized redistribution or identity misuse. Adobe also allows creators to embed a "Do Not Train" signal into the metadata, enabling explicit rejection of inclusion in AI training datasets and contributing to emerging norms of digital rights governance. Once edited, files are saved both locally and in the Adobe Content Credentials Cloud, ensuring continuity of provenance across devices and collaborative workflows. For downstream users such as journalists, fact-checkers, and forensic analysts, Adobe provides a dedicated Inspector Tool, enabling independent verification of authenticity by visualizing the full transformation history of an asset as recorded in the C2PA manifest [32]. For web-based review, Adobe recommends the Chrome Content Credentials Extension, which allows users to inspect provenance directly in browser environments and trace the complete lineage of an image. Taken together, Adobe's software ecosystem illustrates the essential role of post-processing tools in operationalizing C2PA's provenance model across the full lifecycle of digital imagery. By tightly integrating secure, tamper-evident Content Credentials with AI-enhanced editing capabilities, Adobe ensures that authenticity signals originating at capture remain intact, verifiable, and meaningful throughout professional editing, publication, and archival workflows. Figure 3 shows an example for such a workflow.

Beyond Adobe's central role in operationalizing Content Credentials across creative pipelines, several additional software vendors contribute significantly to the emerging ecosystem of provenance-aware tools. **Google** has taken a leading position in the detection and labeling of AI-generated media through SynthID, a watermarking and provenance-support technology developed by Google DeepMind [33]. SynthID embeds imperceptible, machine-readable watermarks directly into generated images and audio, enabling downstream detection even after editing, compression, or platform resharing. While not a full implementation of the C2PA standard, SynthID has been promoted as a complementary approach to authenticity signaling in web-scale environments, supporting cross-platform provenance in alignment with transparency obligations under the AI Act [34]. **Microsoft** has introduced provenance-oriented features through tools such as Microsoft Designer, Bing Image Creator, and the Microsoft Edge Content Credentials Integration, each of which implements or displays C2PA-compatible metadata whenever available [35]. Microsoft collaborates closely with the Content Authenticity Initiative and integrates Content Credentials into its image-generation services, ensuring that AI-created outputs are automatically labeled with tamper-evident C2PA manifests. These manifests can be inspected in Edge's built-in Content Credentials viewer, enabling users, journalists, and digital-forensics analysts to confirm the origin of AI-generated visuals. Microsoft's approach emphasizes interoperability at browser level, making provenance inspection widely accessible. Similarly, **OpenAI** implements provenance and safety metadata within its generative systems, including DALL·E, Sora, and image APIs. OpenAI now attaches C2PA-conformant Content Credentials to a substantial portion of AI-generated images, embedding crucial metadata such as model identifiers, creation timestamps, and usage disclosures [36]. These credentials persist through file downloads and can be inspected using CAI-compatible tools. At the same time, OpenAI continues to research watermarking and transformation-robust signature mechanisms, aiming to close detection gaps exposed by image cropping, filtering, or upscaling operations. The widely use design and content-creation platform **Canva** has integrated Content Credentials into its export workflows to support creators, educators, and marketing teams operating at scale [37]. Canva applies provenance metadata to AI-generated components within composite designs, enabling users to track which elements were produced using automated tools. Through partnerships with the CAI, Canva supports basic C2PA manifest generation and encourages educators and journalists to verify digital assets before publication. Also **Stability AI**, the developer of the Stable Diffusion models, contributes

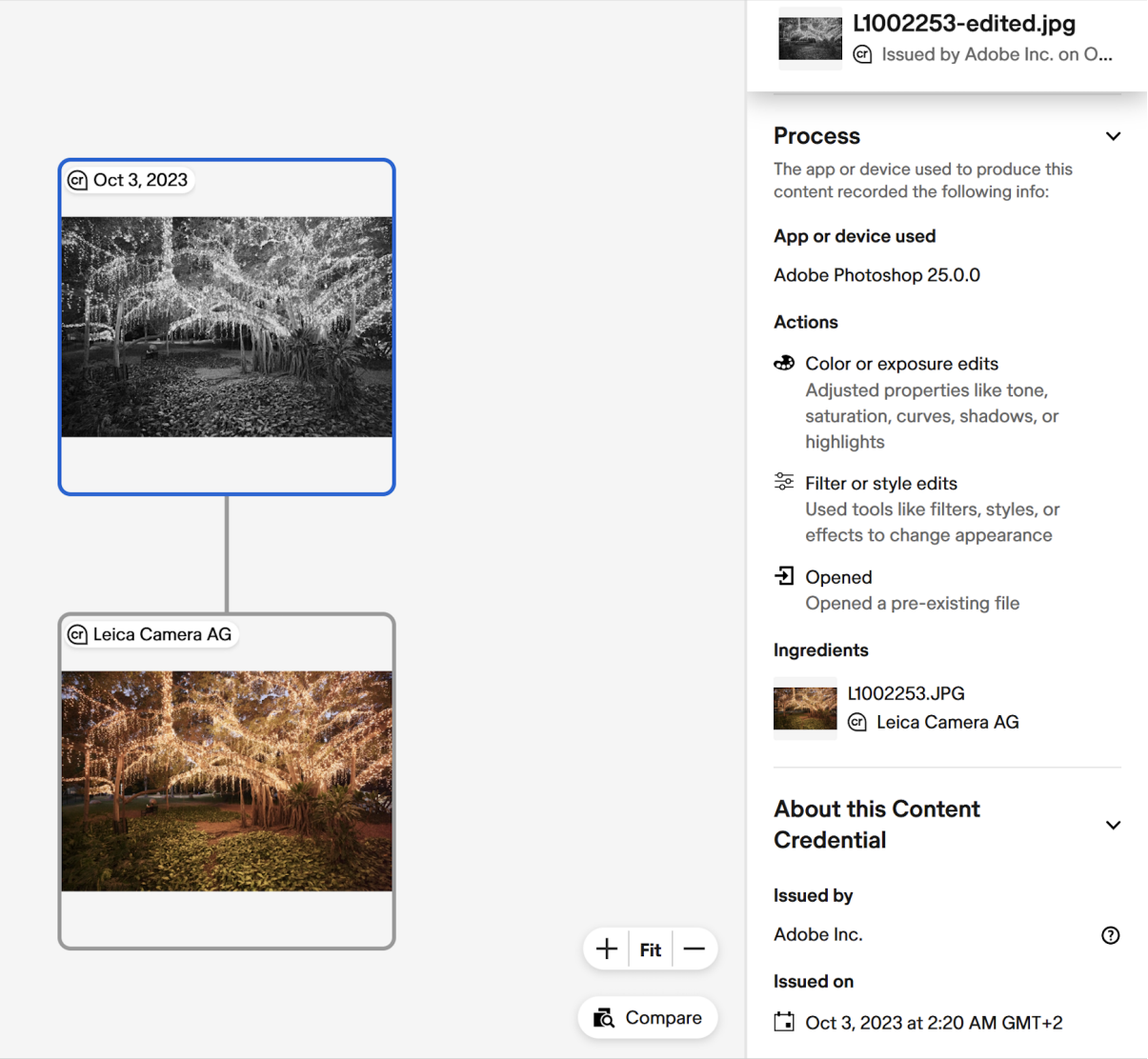

Figure 3: C2PA Content Credentials of an image taken with a Leica camera, edited in Adobe Photoshop.

to provenance efforts by providing metadata-retention pathways and supporting external authenticity layers. While Stable Diffusion does not enforce C2PA manifests natively, Stability AI has integrated structured metadata fields indicating the model version, sampler, and generation parameters. These fields—while not cryptographically bound—support traceability within research and content-creation communities. Stability AI has also expressed public support for emerging industry standards, positioning their ecosystem for future compatibility with C2PA-compliant provenance layers.

Taken together, various software vendors demonstrate the rapid expansion of provenance-aware tooling throughout the digital-content lifecycle. While their technical implementations differ—ranging from watermarking to full C2PA integration—they collectively contribute to a broader infrastructure for authenticity and transparency across generative and editorial workflows.

In alignment with the implementation of C2PA, the successful implementation of eIDAS 2.0 depends not only on regulatory mandates but also on the breadth of industry adoption across identity providers, Qualified Trust Service Providers (QTSPs), enterprise vendors, and device ecosystems. Within the European Union, a broad range of QTSPs—including **D-Trust (Bundesdruckerei Group), Entrust, InfoCert, LuxTrust**, and **SK ID Solutions** already deliver the qualified trust services required under the revised regulation, such as Qualified Electronic Signatures (QES), Qualified Electronic Seals (QSeal),

and Qualified Electronic Timestamps (QTST) [38, 39, 40]. These providers not only ensure compliance with ETSI standards for cryptographic validation but also offer cloud-based APIs and remote-signing infrastructures, enabling seamless integration of eIDAS-compliant signatures into digital workflows used in journalism, public administration, and high-assurance enterprise environments. A pivotal area of industry engagement concerns the emerging European Digital Identity Wallet (EUDI Wallet) ecosystem. Major technology vendors, member-state agencies, financial institutions, and telecom operators have begun piloting PID (Person Identification Data) credentials and Qualified Electronic Attestations of Attributes (QEAAs) according to the EU's Architecture and Reference Framework (ARF) [41]. These pilots lay the groundwork for cross-border interoperability, allowing users to securely present identity and role information in digital transactions—an essential element when linking human actors to C2PA-verified media. Several Member States, including Germany, France, Italy, the Netherlands, and Finland, have initiated **national EUDI Wallet programs** supported by private-sector consortium partners who provide secure wallet components, identity verification modules, and certification-compliant onboarding pipelines [42].

Beyond identity providers, enterprise cloud platforms have substantially expanded support for eIDAS 2.0. Companies such as **Microsoft, Adobe, DocuSign**, and **Signicat** have introduced comprehensive, ETSI-aligned signature stacks that support QES/QSeal creation and validation across business services and content workflows [43, 31]. These services integrate directly with document-management systems, legal processes, and digital-evidence management platforms, ensuring that digital artefacts—whether media files, reports, or legal documents—can be cryptographically verified against the EU Trusted Lists (EUTL). As organizations modernize their internal and cross-border digital processes, eIDAS 2.0-compliant verification has rapidly become a requirement for interoperability in procurement, compliance, and public-sector service delivery. Finally, hardware and device vendors have started adding secure-element and trusted-execution-environment capabilities that support on-device signature creation. Smartphones and professional devices increasingly integrate hardware-based key protection and secure enclave features aligned with eIDAS requirements for "sole control" during signature operations. These developments strengthen the security of remote and mobile QES processes and align with the long-term regulatory objective of enabling users to securely sign digital content—including media assets—directly on consumer devices.

Taken together, industry support for eIDAS 2.0 signals the emergence of a cohesive ecosystem in which legal trust, digital identity, and authenticated media provenance converge. As QTSPs, wallet providers, enterprise platforms, and device manufacturers increasingly align with the regulation, eIDAS 2.0 evolves into a foundational infrastructure for legally robust, interoperable, and tamper-evident digital workflows across the European Union.

2.5.3. Supporting Platforms

Alongside hardware and software vendors, digital platforms increasingly play a decisive role in surfacing, displaying, and enforcing Content Credentials in everyday media consumption. **LinkedIn** is among the earliest large-scale social platforms to provide native support for C2PA-based provenance labeling. Since 2023, LinkedIn has progressively rolled out a multi-phase implementation of Content Credentials indicators, enabling selected user groups to display authenticity signals directly on uploaded images as shown in Figure 4 [44]. The platform uses the standardized Content Credentials icon, ensuring visual consistency with the C2PA user-experience guidelines. By interacting with this symbol, users can inspect essential provenance metadata, such as whether AI tools were involved in the creation, who the verified creator is, and when the asset was originally published. Through this approach, LinkedIn aims to reduce misinformation risks on professional networks by highlighting the origin and editing history of visual assets. Google, beyond its role as a global search and advertising provider, has taken a systemic position within the C2PA ecosystem. In 2024, Google joined the C2PA Steering Committee, becoming a formal governance participant in shaping the technical development of provenance standards [34]. As part of this commitment, Google integrated the visibility of C2PA metadata into its Search and Google

Ads surfaces: users can now inspect provenance information via the “About This Image” interface, which reveals whether an asset was generated with AI, where it first appeared, and which C2PA credentials were embedded by the creator or generating system [33]. These measures substantially increase transparency for billions of users and extend the visibility of provenance data into the advertising domain, where manipulated imagery has historically been difficult to detect. Together, LinkedIn and Google illustrate a crucial shift: provenance is no longer limited to creators and editors, but is becoming an integral part of content consumption platforms, shaping user trust at the point of interaction. Their support for C2PA significantly expands the reach of authenticity signals, helping ensure that provenance metadata is not only generated during creation but also remains visible and meaningful at the moment of public interpretation.

Figure 4: LinkedIn supports C2PA content credentials (cr-symbol on the left, and details on the right).

Beyond search and professional networks, large-scale social and media platforms increasingly determine whether provenance signals are actually visible to everyday users. **TikTok** has begun rolling out C2PA-aligned Content Credentials labeling for eligible accounts and AI-assisted assets, disclosing when generative tools are used and exposing provenance details to viewers through an in-app inspector. The company describes a phased implementation approach with creator tooling, moderation hooks, and downstream display that aligns with C2PA’s disclosure principles [45]. By surfacing whether content was AI-generated and when it was first published, TikTok aims to reduce ambiguity around altered or synthetic media at the point of consumption. Likewise, also **Meta** has expanded transparency for visual media across Facebook and Instagram, combining native “Made with AI” indicators, watermarking research, and Content Credentials display for participating partners [46]. In practice, Meta’s approach combines platform-level labels with support for C2PA manifests where available, so that viewers can inspect core provenance fields (e.g., creator attribution, tool involvement, and publication context), while creators preserve tamper-evident metadata through export and re-share workflows consistent with C2PA guidelines. On **X (formerly Twitter)**, the platform has emphasized contextual transparency rather than full C2PA creation tooling. X has introduced media-focused Community Notes and policy commitments around manipulated media, while experimenting with Content Credentials display where compatible metadata is present [46]. Although not yet a complete C2PA authoring ecosystem, X’s emphasis on viewer-side inspection and public context complements provenance by enabling fast crowd-validated scrutiny of images and videos circulating at scale. Finally, stock-media ecosystems have become pivotal carriers of provenance. **Getty Images** and **Shutterstock** have announced Content Credentials support in their professional pipelines—both for AI-generated assets and for editorial/licensed photography—so that buyers, newsrooms, and rights-holders can verify origin, edit history, and license context on download and reuse. In effect, stock libraries function as provenance concentrators: they co-bundle rights metadata and C2PA manifests, improving trust at the point where assets enter

broadcast, advertising, and newsroom workflows.

2.6. Summary

In this section, we showed that trustworthy multimedia evidence relies on the combined strength of legal, technical, and industrial frameworks. eIDAS 2.0 establishes the legal foundation by providing EU-wide mechanisms for secure digital identity, qualified signatures, seals, and timestamps, ensuring that digital evidence can be authenticated and recognized across borders. C2PA complements this by delivering a technical standard for embedding tamper-evident provenance metadata directly into media files, enabling transparent documentation of how an asset was created, edited, and transmitted. The EU AI Act adds a regulatory layer that enforces transparency for AI-generated and AI-modified content, ensuring that synthetic media can be reliably identified and contextualized. Together, these frameworks create a cohesive environment in which authenticity, integrity, and transparency can be enforced both technically and legally.

The section also demonstrates that research in multimedia retrieval, provenance, and detection technologies forms a strong academic foundation for these developments, while industry adoption across camera manufacturers, software providers, and digital platforms shows that provenance and authenticity mechanisms are moving rapidly into real-world production workflows. Across the entire ecosystem—legal regulations, technical standards, scientific research, and practical implementations—a consistent trend emerges: trustworthy multimedia processing is becoming a central requirement for handling digital evidence, and a wide range of actors are aligning their technologies and practices accordingly. Therefore, in the next section we will provide a detailed modeling for the integration of such services into a wide variety of applications.

3. Modeling

Verification mechanisms, like e.g. C2PA or eIDAS 2.0 need to be incorporated and integrated by Multimedia Information Retrieval (MMIR) systems. As almost any system is working with Multimedia in various aspects, all these systems facilitate a MMIR component for processing and displaying Multimedia in some extent. Therefore, the presented modeling in this section refers to the MMIR reference architecture (see section 2), which provides a model for all relevant aspects of both MMIR processing and AI integration. This reference framework is used to show a generic approach, which is adaptable for any other MMIR-based solution. Basically, making existing MMIR solutions to support verifiable media processing requires two extensions: one on the pre-processing side of MMIR systems, the other on the UI side, where results are presented and users interact with Multimedia. Both aspects transform MMIR systems into Trustworthy and trusted MMIR systems and will now be modeled in this section. Based on the the MMIR reference architecture shown in section 2, Figure 1, these aspects address both the *Submission* and the *Perception* phases as shown in Figure 5.

While large parts of MMIR and AI processing remain untouched, when introducing trust and trustworthiness, particularly, the areas of *Feature Extraction* and *Result Presentation* must be extended. Therefore, in the next subsections, more details of these extensions are presented.

3.1. Pre-Processing

In general, every Multimedia system has a component responsible for capturing (i.e., creating), ingesting (i.e., importing) Multimedia assets, such as images, text, video, audio, or combined formats. Typically, some kind of pre-processing is required to decide, which importing facilities need to be employed, or which feature extraction mechanisms need to be applied to the media asset. This *Feature Extraction* phase can both be required for submissions to the MMIR system, and for the processing of AI generated content as part of a MMIR system. In both cases, content credentials can be attached to the original assets and have to be included in the feature extraction and further processing phases. Typically, MMIR systems employ import filters for the extraction of various features. Therefore, a trustworthy extension

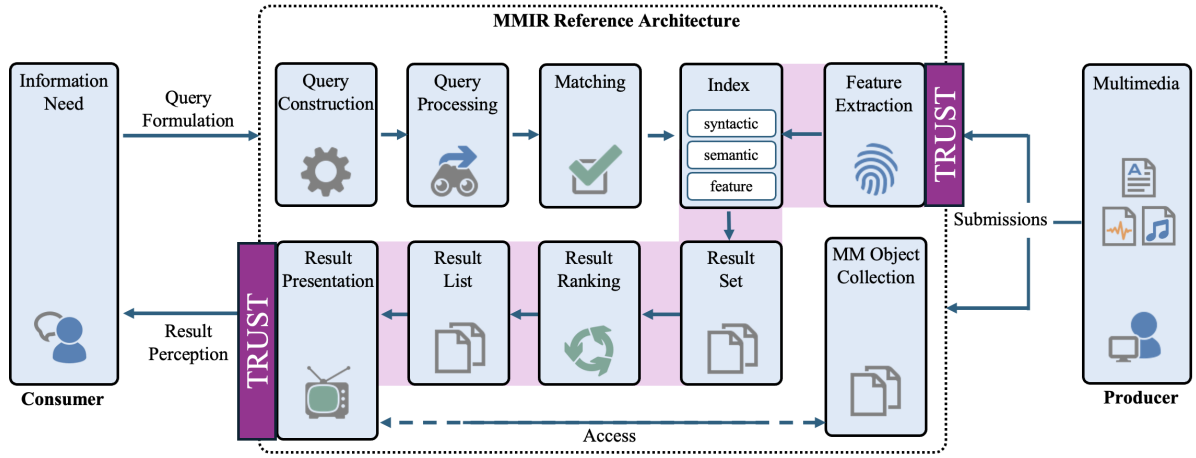

Figure 5: Trust at the system borders of a MMIR system.

to such frameworks can be modeled by integrating an additional layer of filters into their standard configuration:

A *TW-Filter* is an abstract structure encapsulating all filters in the context of *Trustworthiness*. It has three subclasses representing the major areas, where trust can be established: the *Watermark*-Filter, which is responsible for detecting watermarks of Multimedia content denoting their origin - including compliance to the EU AI-Act's transparency demands; the *Fake-Detection*-Filter, which represents a collection of available mechanisms to detect, if ingested Multimedia is generated or manipulated; and the *Verification*-Filter representing the parent-structure for extendable legal or industry verification standards, like the C2PA or eIDAS 2.0 as presented in this paper.

All *Trustworthiness-Filters* have a similar structure with the following elements:

- **confidence**: a value between 0 and 1 denoting the confidence of the detection algorithm. For *Watermark-* and *Verification*-Filters this will be either 1 (100% confident) or 0 (not confident at all), while for other filters, like e.g. the *Fake-Detection*, any value between 0 and 1 can be returned.
- **icon**: an icon representing the pre-processing method. In case of C2PA or eIDAS 2.0, the standard icons of the corresponding initiatives will be returned.
- **description**: each pre-processing filter can extract various information depending on its purpose. These can be certificate data or expiration dates, editing history, or other metadata. To support maximum flexibility, for the description a JSON-format is employed.

The results of the pre-processing must be regarded as Multimedia features and permanently attached to the processed Multimedia asset. The GMAF, e.g., uses a *Multimedia Feature Graph (MMFG)* to represent all extracted features of a Multimedia asset. As the MMFG already is a generic structure, it can simply be extended by adding an additional node named *Trustworthiness*, where all results of any pre-processing Filter are added in a textual representation. This direct integration with the MMFG ensures that any processing or display of Multimedia assets can access this information.

3.2. Presentation and Interaction

On the left side of Figure 5, the Users' perception is shown, which has to be extended in two regards to meet the requirements of C2PA and eIDAS, respectively of trust in general: first, the *result presentation* must be extended to show the trustworthiness of every Multimedia asset in form of a simple icon, indicating the pre-processing status. Whenever an asset has been verified, watermarked, or detected as being fake, a symbol informs the user. Second, an additional *interaction* is required to give users access to the processing details, i.e., the description of the corresponding Trustworthiness-Filter. Figure 6

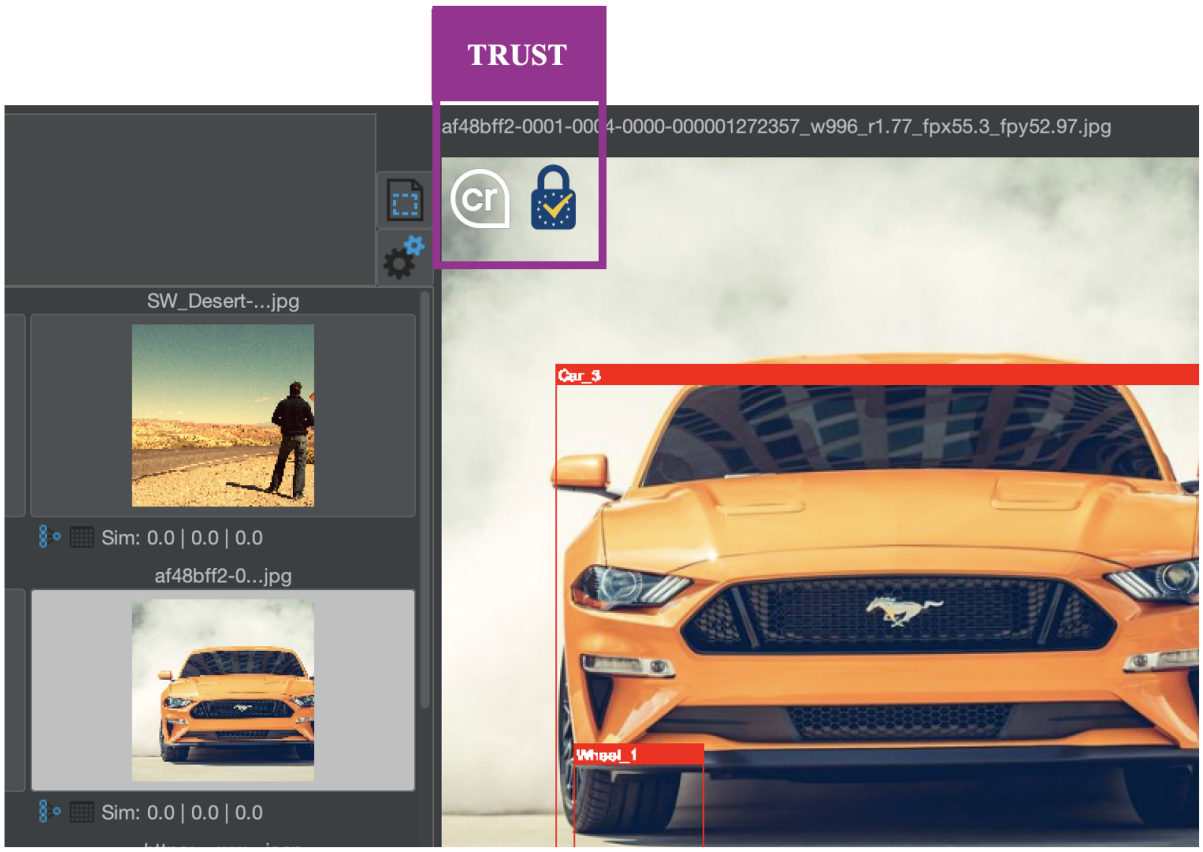

Figure 6: Presenting trust-information to the users in a MMIR system, exemplary shown with the GMAF framework [17].

shows this in more detail.

A framework, which complies to both trusted pre-processing and the corresponding transparent presentation of and interaction with result elements, can be regarded as trustworthy MMIR framework.

3.3. Conflicts, Rules and Conditions

In practical forensic and legal workflows, verification mechanisms may yield conflicting trust signals. In particular, content-centric provenance mechanisms such as C2PA and actor-centric legal trust services under eIDAS 2.0 address orthogonal aspects of trust and therefore must not be collapsed into a single confidence value. C2PA expresses statements about the integrity and lifecycle of a media object ("what happened to the content"), while eIDAS 2.0 establishes legally qualified attribution and intent ("who acted, when, and under which legal mandate"). Consequently, a valid result in one dimension does not compensate for failure in the other.

To address this systematically, the *Trustworthiness-Filter* evaluates the consistency of extracted trust signals using deterministic rules. Typical conflict scenarios include files that carry a valid qualified electronic signature but whose C2PA provenance chain is broken, as well as technically intact assets lacking legally qualified attribution. These situations are not treated as errors but as *semantically meaningful outcomes* that directly affect evidentiary weight. At presentation level, MMIR systems communicate such conflicts through layered interaction principles: a primary icon-based status representation, a conflict indicator when applicable, and an explicit natural-language interpretation. This approach avoids opaque trust scores, ensures explainability, and aligns with judicial expectations regarding transparency and burden-of-proof assessment.

3.4. Further Considerations

Conceptual models such as the MMIR provide a reference architecture for integrating AI, multimedia, and trust mechanisms. It extends traditional MMIR systems with modules for provenance, identity management, and legal compliance, enabling end-to-end workflows for forensic and judicial applications (see Figure 5). Conceptually, MMIR provides the socio-technical backbone into which C2PA and eIDAS 2.0 are integrated as complementary assurance layers. C2PA contributes verifiable *content provenance* at the media layer, while eIDAS 2.0 supplies *legal trust* through qualified signatures, seals, time-stamps, and attribute attestations. The goal is a single chain of custody whose artefacts are, at once, machine-verifiable (C2PA manifests, bindings, and validation semantics) and legally recognised across the EU (eIDAS trust services and wallet-based identity/role proofs), without breaking MMIR's modularity or evolution paths [4, 5, 47]. This conceptual alignment can be summarized as follows: C2PA attaches *what happened to the content, when, and how* at the multimedia layer; eIDAS attaches *who did it, under what legal mandate, and with what probative guarantees* at the trust layer; and MMIR orchestrates these layers into a coherent pipeline that begins at capture / ingest and ends in verifiable disclosure and presentation, while remaining modular enough to accommodate evolving assertion vocabularies, wallet formats, and supervisory requirements [4, 5].

3.5. Summary

In this section, the modeling for a trustworthy extension to existing MMIR solutions has been presented. Both on capture/ingest and on the presentation side of MMIR solutions, mechanisms must be integrated to ensure a transparent, reliable handling of verification and authenticity information. Users are informed about the content credentials and trustworthiness of specific assets by symbols in the user interface, and the MMIR solution incorporates import filters to extract such information and append it to the standard datastructures within the MMIR system. Based on the GMAF, these extensions are modeled, resulting in a trustworthy MMIR reference architecture. In the next section, further insights on the prototypical implementation are given.

4. Prototypical Implementation

In this section, insights about the prototypical implementation of the integration of C2PA and eIDAS 2.0 into the GMAF are given. As shown in the previous section, two major areas of implementation need to be considered: the user interface, showing verification symbols and opening a metadata dialogue; and the pre-processing, where the actual verification of Multimedia must be implemented. The first area is quite straight-forward and based on the verification results of the pre-processing. Therefore, and according to publishing constraints, in this section the second area will be presented in more details. First, we will outline the integration of C2PA in subsection 4.1, followed by the integration of eIDAS 2.0 in subsection 4.2.

4.1. Technical Integration of C2PA

From a technical and implementation perspective, a production-grade solution starts with a clear understanding of the C2PA data model and its on-asset representation. Every C2PA-enabled media file carries a manifest store comprising one or more manifests. The active manifest is the one a validator must locate and verify. Each manifest contains a claim, which is a digitally signed structure enumerating assertions that describe creation, editing, composition, and related provenance facts. Assertions may be embedded or externally referenced and can carry JSON or binary payloads, whose schema, multiplicity, and redaction rules are defined at specification level. For on-asset packaging, C2PA relies on *JPEG Universal Metadata Box Format (JUMBF)* containers to store and reference manifest components inside media binaries (e.g., JPEG, PNG, BMFF) so that the claim, its signature, and assertion set are bound to the asset in a tamper-evident way [4].

Binding the claim to the content is central to authenticity. C2PA supports hard bindings, which hash specific pixel regions or container boxes and include those digests in the claim; any alteration of the bound regions invalidates verification. It also supports soft bindings, which introduce recoverable identifiers (fingerprints or invisible watermarks) that allow validators to resolve a manifest even when it has been stripped during transcoding or platform processing. The specification defines the semantics for both binding modes, while the ecosystem exposes cloud interfaces to look up soft-bound manifests when needed [4, 48].

Cryptographic trust and identity follow familiar public-key patterns. Claims are signed using X.509 certificates, validated against trust lists, with explicit checks for time-stamps (via a TSA) and revocation (CRLs/OCSP). A standard-compliant validation thus performs a deterministic pipeline: identify the active manifest, locate and parse the claim, verify the signature and time-stamp, confirm certificate validity or revocation status, validate each assertion (including cross-references and hashes), and recompute content bindings to ensure the asset still matches the claim. The specification also prescribes how results should be presented at the user interface level, including disclosure principles that underpin the familiar Content Credentials pin [4, 49]. Developers typically build on the CAI open-source SDK, which provides a Rust reference implementation (c2pa-rs) and bindings for JavaScript/TypeScript (Node and browser), Python, C/C++, Java, and mobile platforms. The SDK supports creation and signing of claims/manifests, embedding in supported formats, and validation logic consistent with the specification.

In our prototypical implementation we therefore used a command-line-wrapper to implement the *Trustworthiness-Filter* mentionend in the modeling section of this paper. It basically executes system commands, receives and interprets the results. Therefore, in the following listing, the original command and toolset is described instead of the wrapper-implementation. Due to the existing extension points of MMIR, as implementation for the extraction of C2PA verification data, the *c2patool* has been integrated, as it provides a production-ready, pre-configured environment for many platforms. To extract the JSON metadata from a C2PA verified Multimedia asset, the code snippet shown in listing 1 is used.

Listing 1: Extracting C2PA Content Credentials from a file.

```
// Initiate the extraction of C2PA data via command line
./c2patool MyImage.jpeg

// Resulting JSON Object (extract)
{
"active_manifest":
"urn:c2pa:36a5bc33-ac27-4c22-bfb0-0b9720b9a5bf:adobe",
"manifests": {
"adobe:urn:uuid:7409484f-6826-45b8-ae91-a9ff4b8bdb5a": {
  "claim_generator": "Adobe_Stock c2pa-rs/0.4.2",
  "title": "AdobeStock_94253216.jpeg",
  "format": "image/jpeg",
  ...

  "claim_generator_info": [
  {
    "name": "Adobe Content Authenticity",
    "com.adobe.aca-version": "a821da7",
    "org.contentauth.c2pa_rs": "0.67.1"
  }
  ],
  "title": "AdobeStock_94253216.jpeg",

  ...

  "validation_results": {
  "activeManifest": {
    "success": [
    {
      "code": "signingCredential.ocsp.notRevoked",
      "url": "OCSP_RESPONSE",
```

```
32        "explanation": "certificate not revoked"
33      },
34      {
35        "code": "timeStamp.validated",
36        "url": "self#jumbf=/c2pa/adobe:urn:uuid:
37        7409484f-6826-45b8-ae91-a9ff4b8bdb5a/c2pa.signature",
38        "explanation": "timestamp message digest matched:
39        DigiCert Timestamp 2022 - 2"
40      },
41
42 ...
```

While the initial call of *c2patool* is quite short (see listing 1, line 2), the interpretation of the result is more complicated. The most important sections of the resulting JSON document are shown in listing 1: lines 5-12 and 15-22 show, that the file has been verified by *Adobe Content Authenticity*. It also shows the original file name (line 11 and 22), which deviates from the name *MyImage.jpeg* used when calling the tool. Furthermore, the validation results are shown in lines 26-40 indicating a successful verification. The complete JSON can be used to reveal the active manifest, claim generator, assertions (e.g., c2pa.actions, stds.schema-org.CreativeWork), and signature_info—or, in detailed mode, assertion hashes and JUMBF URLs that correspond to the binary embedding [50, 51]. Depending on the use cases of the MMIR solution, such metadata can be relevant for further processing or presentation. However, in the context of the reference framework, a detailed further use can be omitted.

4.2. Technical Integration of eIDAS 2.0

Also, for the implementation of eIDAS 2.0 services, we selected a command-line based approach making the integration with the MMIR both simple and flexible. There are several provides for eIDAS 2.0 verification services, which are listed by the European Union at [52]. For our implementation, we used the European Union's DSS (Digital Services) Open Source library [53, 54].

Under eIDAS 2.0, qualified trust services provide legal presumptions of authenticity, integrity, and time for digital artifacts. As our goal is to make multimedia evidence legally robust across the EU, we implemented a Java class compliant to the GMAF based on the DSS library. Listing 2 shows the details.

Listing 2: Extracting eIDAS 2.0 information from a file.

```
1 import eu.europa.esig.dss.model.*;
2 import eu.europa.esig.dss.validation.*;
3 import eu.europa.esig.dss.simplereport.*;
4
5 public class EidasValidator extends TrustworthinessFilter {
6   ...
7   public void validate(File f) {
8     Document document = new FileDocument(f);
9     SignedDocumentValidator validator =
10    SignedDocumentValidator.fromDocument(document);
11    Reports reports = validator.validateDocument();
12
13    System.out.println(
14    reports.getSimpleReport().getValidationIndication());
15  }
16 }
```

Basically, this listing illustrates, that the verification process is quite simple. In line 8, a Multimedia file is handed over to the DSS library's *Document* object. With this, a *Validator* can be created (lines 9-10), which then gives access to reporting facilities. Such a report can be formatted as XML and contains the data shown in Listing 3

Listing 3: Sample Report for an eIDAS 2.0 signature process.

```
1 <SimpleReport>
2   <ValidationPolicy>QES_POLICY_V1</ValidationPolicy>
```

```
3   <DocumentName>vertrag.pdf</DocumentName>
4   <ValidationTime>2026-01-03T10:58:11Z</ValidationTime>
5   <OverallIndication>VALID</OverallIndication>
6
7   <Signature>
8     <Id>sig-1</Id>
9     <SignedBy>Dr. Max Mustermann</SignedBy>
10    <SignatureFormat>PAdES_BASELINE_T</SignatureFormat>
11    <SignatureQualification>QES</SignatureQualification>
12    <Indication>VALID</Indication>
13    <SigningTime>2025-12-22T14:03:11Z</SigningTime>
14
15    <CertificateChain>
16    <Certificate
17      subjectCN="Dr. Max Mustermann"
18      issuerCN="QTSP Qualified CA" />
19    <Certificate
20      subjectCN="QTSP Qualified CA"
21      issuerCN="EU Root" />
22    </CertificateChain>
23    <Timestamps>
24    <Timestamp type="SIGNATURE_TIMESTAMP"
25      indication="VALID"
26      productionTime="2025-12-22T14:03:15Z"
27      tsa="Qualified TSA Name"/>
28    </Timestamps>
29  </Signature>
30
31  <Countersignatures>0</Countersignatures>
32  <NumberOfSignatures>1</NumberOfSignatures>
33 </SimpleReport>
```

The simple report shown in Listing 3 already illustrates the provided information for verification purposes. In line 3, the validated document, and in line 4 the validation time is contained. The overall indication (line 5) says, the document is *valid*. Furthermore, the signature details (lines 7-13) and the certificate chain (lines 15-23) is available. With this information, the chain of signatures, certificates, and timestamps of a document can be verified in a simple and flexible way.

In summary, the state of the art in trustworthy multimedia processing is characterized by the convergence of advanced AI techniques, interoperable provenance standards, and evolving legal frameworks. Ongoing research continues to address open challenges in robustness, explainability, privacy, and cross-border interoperability, with a growing emphasis on practical implementation and empirical evaluation.

5. Evaluation

The prototype demonstrates that a MMIR aligned stack can integrate C2PA's tamper-evident provenance with eIDAS 2.0's legally binding trust services in a coherent end-to-end chain of custody. On the provenance side, manifests, claims, and typed assertions are embedded via JUMBF and cryptographically bound to the asset; validation traverses signer identity, trust lists, revocation, and content bindings as prescribed by the C2PA v2.2 specification, including disclosure principles for user-facing inspection of Content Credentials [4, 49]. On the legal trust side, qualified signatures, seals, and time-stamps are created and validated with the EU Digital Signature Services (DSS) library in accordance with eIDAS 2.0, with ETSI-conformant reports to support courtroom disclosure and continuous alignment to Trusted List v6 (TLv6) migration requirements [53, 55]. Within the MMIR system, these pieces attach cleanly to the Trust Layer (CCI) for cryptographic and policy validation, the Identity & Attribute plugin for wallet interactions, the QTSP orchestration for remote QES/QSeal when required, and the Audit/Archive

module for durable preservation of artifacts (signature values, TST tokens, trust-chain references, and ETSI reports) [7, 54].

Empirically, the integration achieves its principal objectives: (i) provenance data remains machine-verifiable across capture, edit, and distribution, even when platforms strip metadata, since soft binding enables manifest recovery via fingerprint/watermark resolution; (ii) legal presumptions of authenticity, integrity, and time are attached to the same assets through qualified signatures and time-stamps; (iii) identity and roles are linked to actions and submissions through PID/(Q)EAA credentials; and (iv) verification results are exportable in ETSI-compliant reports, facilitating admissibility and audit trails in legal proceedings [48, 4, 53]. The MMIR reference architecture proves beneficial: provenance and trust services are treated as composable plugins sourced from standard interfaces, enabling independent evolution (e.g., C2PA assertion catalog growth or wallet protocol updates) while maintaining coherent validation semantics and evidentiary readiness.

At the same time, the evaluation reveals attack surfaces and systemic weaknesses that call for operational hardening. Content bindings are effective against direct tampering of authenticated regions or boxes, but adversaries may attempt *transcode-stripping attacks* by re-encoding media to remove JUMBF containers and thereby detach on-asset manifests; soft binding mitigates this via watermark or fingerprint lookup, yet becomes dependent on resolution services and approved algorithms, which introduces availability and side-channel risks if those services are degraded or if fingerprints collide in edge cases [48, 4]. *Certificate and trust-list freshness attacks* exploit stale TLv6 data or revocation gaps: validators must fail closed when LOTL/EUTL cannot be refreshed, and audit trails must evidence the freshness window; otherwise, signatures could be incorrectly accepted under outdated trust conditions [55, 53]. *Time-stamp replay or backdating* is prevented by qualified TSA policies, but if the TSA is unreachable or misconfigured, systems may defer time anchoring or fall back to non-qualified marks; strict enforcement of TSA availability and policy requirements is essential to preserve probative value [53, 47]. In wallet flows, *attribute inflation or over-disclosure* can undermine privacy and increase legal exposure; evaluators should monitor selective disclosure ratios and reject presentations exceeding the requested set, thereby adhering to the Rulebook's minimization principles [56, 57]. For remote QES, *QTSP throttling or cross-device interception* is a realistic operational risk; CSC endpoints must enforce mutual authentication, channel binding, and nonce-based challenge-responses, with clear recovery strategies when the QTSP degrades, to avoid signature delays or incomplete sealing of submissions [58, 47]. Finally, *provenance spoofing via bogus ingredient graphs*—where malicious actors embed deceptive references to third-party assets—requires rigorous validation of ingredient manifests and their chains, not merely superficial presence; the C2PA validation process should reject unverifiable or inconsistent ingredient claims and flag them in the ETSI reports and viewer UX [4, 49].

Overall, the prototype meets the evaluation goal of demonstrating that the MMIR reference architecture and framework can unify provenance, identity, and legal trust into a single pipeline. Its strengths lie in standards-based interoperability and layered validation, while residual vulnerabilities appear primarily in the operational envelope: transcode durability, trust-list currency, TSA/QTSP availability, and privacy discipline in wallet-based disclosures.

6. Discussion and Summary

While the proposed integration of C2PA, eIDAS 2.0, and the MMIR reference architecture demonstrates the technical and architectural feasibility of end-to-end trustworthy multimedia processing, several limitations must be acknowledged and provide valuable directions for future research.

First, the model deliberately refrains from collapsing heterogeneous trust signals into a single numerical score or automated evidentiary decision. C2PA and eIDAS address fundamentally different trust dimensions—content integrity and legal attribution respectively—and their outcomes are therefore modeled as orthogonal rather than compensatory. While this avoids opaque or legally questionable trust aggregation, it also implies that the framework itself does not determine evidentiary weight. Assessing probative value remains a normative task dependent on legal domain, jurisdiction, and procedural

context. Future work may therefore explore configurable evidence profiles that map verifiable trust states to domain-specific recommendations without undermining transparency or judicial discretion.

Second, the robustness of the approach depends on external trust infrastructures. C2PA soft-binding mechanisms rely on resolution services, while eIDAS validation presupposes timely access to trusted lists, revocation data, and qualified trust service providers. Although these dependencies are inherent to any standards-based trust ecosystem, they raise questions about availability, governance, and long-term verifiability. Further research is required on archival strategies that preserve not only signed media assets but also the contextual trust material (e.g., trust-list snapshots, OCSP responses) necessary for retrospective verification years after submission.

Third, the evaluation presented in this paper focuses on architectural integration and technical correctness rather than large-scale empirical validation. The usability and interpretability of conflict indicators—such as cases where legally valid signatures coexist with broken provenance chains—have not yet been studied with real users. Future empirical work involving judges, forensic experts, journalists, or legal practitioners could investigate whether the proposed presentation mechanisms effectively support correct trust assessment without introducing cognitive bias.

Finally, while the model assumes compliance with privacy-preserving wallet principles under eIDAS 2.0, fine-grained governance of attribute disclosure remains an open challenge. Over-disclosure of identity or role attributes may increase legal exposure or conflict with data-minimization principles. Extending the framework toward privacy-adaptive trust views—where different stakeholders receive tailored, minimal representations of trust information—represents an important avenue for future work.

In summary, the proposed framework establishes a solid, standards-based foundation for trustworthy multimedia evidence, while leaving intentionally open those normative, organizational, and empirical questions that must be addressed in close collaboration between technical, legal, and institutional actors.

Author Contributions

All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Stefan Wagenpfeil, Joyce Amy Kolenda, Waldemar Nauvomets, Tim Ternovik, and Jakob Wenzig. The first draft of the manuscript was written by Stefan Wagenpfeil and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Declaration on Generative AI

During the preparation of this work, the authors used Microsoft Copilot for AI-assisted copy-editing. Microsoft Copilot was used to provide suggestions regarding the clarity and structure of manually selected sections in order to improve the quality of writing. It was used solely for suggestions on structure, phrasing, and critique. No AI-generated text was copied or directly incorporated into the manuscript. All final wording and scientific content are the authors' own, and the authors take full responsibility for the content of the publication.

References

[1] R. Chesney, D. K. Citron, Deepfakes and the new disinformation war: The coming age of post-truth geopolitics, Foreign Affairs 98 (2019) 147–155. URL: https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war.

[2] D. Kowald, S. Scher, V. Pammer-Schindler, et al., Establishing and evaluating trustworthy ai: overview and research challenges, Frontiers in Big Data 7 (2024).

[3] A. Nastoska, B. Jancheska, M. Rizinski, D. Trajanov, Evaluating trustworthiness in ai: Risks, metrics, and applications across industries, Electronics 14 (2025) 2717.

[4] Coalition for Content Provenance and Authenticity, Content credentials: C2pa technical specification v2.2, 2025. URL: https://spec.c2pa.org/specifications/specifications/2.2/specs/C2PA_Specification.html, technical specification.
[5] European Union, Regulation (eu) 2024/1183 (eidas 2.0), 2024. URL: https://eur-lex.europa.eu/eli/reg/2024/1183/oj/eng, official Journal of the European Union.
[6] European Commission, European digital identity wallet (eudi wallet), 2025. URL: https://commission.europa.eu/topics/digital-economy-and-society/european-digital-identity_en.
[7] S. Wagenpfeil, Towards the trustworthy integration of multimedia and artificial intelligence, 2026. URL: https://www.fernuni-hagen.de, university of Hagen Deposit.
[8] J. F. Nunamaker, M. Chen, T. D. M. Purdin, Systems development in information systems research, Journal of Management Information Systems 7 (1991) 89–106.
[9] E. Union, Artificial intelligence act (regulation (eu) 2024/1689), Official Journal of the European Union, June 2024, 2024. URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32024R1689.
[10] M. Hemmje, S. Wagenpfeil, Multimedia Information Retrieval: Informationen aus multimedialen Quellen gewinnen, Springer, 2025. URL: https://link.springer.com/book/9783662718865.
[11] E. Y. Chang, Foundations of Large-Scale Multimedia Information Management and Retrieval: Mathematics of Perception, Springer, 2011.
[12] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, L. Fei-Fei, Imagenet: A large-scale hierarchical image database, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.
[13] L. Mei, S. Mo, Z. Yang, C. Chen, A survey of multimodal retrieval-augmented generation, 2025. arXiv:2504.08748.
[14] S.-C. Lin, et al., Mm-embed: Universal multimodal retrieval with multimodal llms, 2025. arXiv:2411.02571.
[15] S. Wagenpfeil, B. Vu, P. Mc Kevitt, M. Hemmje, Fast and effective retrieval for large multimedia collections, Big Data and Cognitive Computing 5 (2021) 33. URL: https://www.mdpi.com/2504-2289/5/3/33.
[16] S. Wagenpfeil, P. M. Kevitt, M. Hemmje, Smart multimedia information retrieval, 2022. URL: https://ub-deposit.fernuni-hagen.de/receive/mir_mods_00001858, fernUniversität in Hagen.
[17] S. Wagenpfeil, P. Mc Kevitt, M. Hemmje, Towards automated semantic explainability of multimedia feature graphs, Information 12 (2021) 502. URL: https://www.mdpi.com/2078-2489/12/12/502.
[18] Sony Corporation, Sony delivers highly anticipated firmware updates including c2pa compliancy, 2024. URL: https://www.sony.eu/presscentre/sony-delivers-highly-anticipated-firmware-updates-including-c2pa-compliancy-and-ensuring-authenticity-of-ima press release.
[19] J. Gray, Sony's pxw-z300 camcorder is world's first to support content authenticity, PetaPixel (2025). URL: https://petapixel.com/2025/07/22/sonys-pxw-z300-camcorder-is-worlds-first-to-support-content-authenticity/.
[20] M. Clark, Sony's new camcorder is breaking ground for content authenticity, DPReview (2025). URL: https://www.dpreview.com/articles/5048769642/sony-camcorder-c2pa-content-authenticity-for-video.
[21] Leica Camera AG, Leica m11-p camera with integrated content credentials, 2026. URL: https://leica-camera.com/de-DE/fotografie/kameras/m/m11-p-silber.
[22] Coalition for Content Provenance and Authenticity, C2pa and content credentials explainer (v2.2), 2025. URL: https://spec.c2pa.org/specifications/specifications/2.2/explainer/Explainer.html.
[23] Leica Camera AG, Leica cameras with content credentials (m11-d, sl3-l, m ev1, q3 monochrom), 2026. URL: https://leica-camera.com/de-DE/fotografie/content-credentials?srsltid=AfmBOoqUl9gE70F7PAK6LN6pDJK2hVOONtsfScIt2oGFHUU4KzlLvMKI.
[24] Nikon Corporation, Nikon authenticity service, 2026. URL: https://www.nikonusa.com/content/nikon-authenticity-service.
[25] Nikon Corporation, Firmware update for nikon z6 iii enables c2pa content veri-

fication, 2025. URL: https://www.nikon.de/de_DE/learn-and-explore/magazine/gear/nikon-z6iii-firmware-update-to-feature-content-verification.
[26] M. Kemper, Nikon withdraws c2pa certificate after security flaw in z6 iii, 2025. URL: https://www.photografix-magazin.de/nikon-widerruft-alle-c2pa-zertifikate-nach-sicherheitsluecke-in-z6-iii.
[27] Nikon Corporation, Using c2pa with nikon image cloud (z6 iii online manual), 2026. URL: https://onlinemanual.nikonimglib.com/z6III/de/adding_c2pa_426.html.
[28] Canon Inc., Canon imaging pipeline and metadata integrity information, 2026. URL: https://global.canon/en/technology/imaging/.
[29] Fujifilm Corporation, Fujifilm image processing and transparency initiatives, 2026. URL: https://fujifilm-x.com/global/stories/.
[30] Panasonic Corporation, Panasonic lumix professional video technology overview, 2026. URL: https://panasonic.com/global/consumer/lumix/.
[31] Adobe Inc., Adobe digital signature services and eidas support, 2026. URL: https://helpx.adobe.com/sign/integration/digital-signature-service.html.
[32] Adobe Inc., View and verify content credentials, 2025. URL: https://helpx.adobe.com/creative-cloud/apps/adobe-content-authenticity/content-credentials/view-content-credentials.html, adobe Help Center.
[33] Google DeepMind, Synthid: A tool for watermarking and identifying ai-generated content, 2024. URL: https://deepmind.google/models/synthid/.
[34] Google, How we're increasing transparency for gen ai content with the c2pa, 2024. URL: https://blog.google/innovation-and-ai/products/google-gen-ai-content-transparency-c2pa/.
[35] Microsoft Corporation, Content credentials in azure openai, 2026. URL: https://learn.microsoft.com/en-us/azure/foundry-classic/openai/concepts/content-credentials.
[36] OpenAI, C2pa in chatgpt images, 2025. URL: https://help.openai.com/en/articles/8912793-c2pa-in-chatgpt-images, openAI Help Center.
[37] Canva Pty Ltd, Content credentials adoption, 2025. URL: https://contentcredentials.org/.
[38] European Commission, eidas dashboard – trusted list of qualified trust service providers, 2026. URL: https://eidas.ec.europa.eu/efda.
[39] Entrust Corporation, What is eidas 2? explore today's compliance landscape, 2026. URL: https://www.entrust.com/resources/learn/eidas-2.
[40] Bundesdruckerei Group (D-Trust), eidas 2.0: Our changes at a glance, 2026. URL: https://www.bundesdruckerei.de/en/innovation-hub/eidas/eidas-2-0.
[41] European Commission, European digital identity wallet (eudi wallet), 2026. URL: https://commission.europa.eu/topics/digital-economy-and-society/european-digital-identity_en.
[42] IDnow GmbH, eidas 2.0 and the european digital identity framework, 2026. URL: https://www.idnow.io/blog/eidas-2-0.
[43] Microsoft Corporation, Microsoft cloud services and eidas compliance, 2026. URL: https://learn.microsoft.com/en-us/compliance/regulatory/offering-eu-eidas.
[44] LinkedIn Corporation, Content credentials on linkedin, 2024. URL: https://www.linkedin.com/help/linkedin/answer/a6282984?lang=en-US.
[45] TikTok, Partnering with our industry to advance ai transparency, 2024. URL: https://newsroom.tiktok.com/partnering-with-our-industry-to-advance-ai-transparency-and-literacy.
[46] Content Authenticity Initiative, Industry adoption of content credentials, 2025. URL: https://contentauthenticity.org/blog/.
[47] Entrust, What is eidas 2? explore today's compliance landscape, 2025. URL: https://www.entrust.com/resources/learn/eidas-2.
[48] Adobe, Cai soft binding resolution api, 2025. URL: https://developer.adobe.com/cai-soft-binding-api/.
[49] Content Credentials, Verify media authenticity, 2025. URL: https://contentcredentials.org.
[50] Content Authenticity Initiative, Manifest examples (c2patool simple/detailed), 2024. URL: https://opensource.contentauthenticity.org/docs/manifest/manifest-examples/, updated 2025.
[51] Content Authenticity Initiative, Cai open source sdk, 2025. URL: https://opensource.

contentauthenticity.org/.
[52] European Commission (DIGITAL), eidas dashboard, 2025. URL: https://eidas.ec.europa.eu/efda/trust-services/browse/eidas/tls.
[53] European Commission (DIGITAL), Digital signature service (dss) documentation and demos, 2025. URL: https://ec.europa.eu/digital-building-blocks/DSS/webapp-demo/doc/dss-documentation.html.
[54] esig/dss, Dss: Digital signature service (github repository), 2025. URL: https://github.com/esig/dss.
[55] European Commission (DIGITAL), Trusted list v6 migration notice (etsi ts 119 612 v2.4.1) — applies from 28 april 2026, 2025. URL: https://ec.europa.eu/digital-building-blocks/sites/spaces/DIGITAL/pages/467109107/Digital+Signature+Service+-+DSS.
[56] European Commission / EUDI, Pid rulebook (arf annex 3.01; encodings mso_mdoc and sd-jwt vc), 2025. URL: https://eudi.dev/2.2.0/annexes/annex-3/annex-3.01-pid-rulebook/.
[57] eu-digital-identity wallet, Eudiw issuer: Pid and (q)eaa provider service (oid4vci), 2025. URL: https://github.com/eu-digital-identity-wallet/eudi-srv-web-issuing-eudiw-py.
[58] eu-digital-identity wallet, Wallet-driven & rp-centric remote qes qtsp (csc api), 2025. URL: https://github.com/eu-digital-identity-wallet/eudi-srv-web-walletdriven-rpcentric-signer-qtsp-java.