Tampere University

Baranaba Mugabane

A DIGITAL CONTENT PROVENANCE APPROACH TO SUPPORT ATTRIBUTION, COPYRIGHT, INTELLECTUAL PROPERTY CLAIMS, AND GOVERNANCE FOR HUMAN-AI CREATED WORKS

Master's thesis, triple degree
Faculty of Information Technology and Communication Sciences
Tampere University
Tallinn University
School of Digital Technologies
Lusófona University
Faculty of Natural Sciences, Engineering and Technologies
May 2026

Abstract

Baranaba: A Digital Content Provenance Approach to Support Attribution, Copyright, Intellectual Property Claims, and Governance for Human-AI Created Works
Master's Thesis
Tampere University
Master's Programme in Sustainable Societies and Digitalisation
May 2026

Digital content is increasingly produced through workflows involving generative artificial intelligence, raising complex questions of authorship, copyright, disclosure, licensing, responsibility, and governance. These questions can only be answered meaningfully when reliable information exists about the people, tools, resources, and processes involved in the creation of a work. Yet a persistent informational gap remains. This research addresses that gap by designing and evaluating a governance-oriented provenance approach for human-AI-created works.

The study adopts a Design Science Research Methodology and combines normative doctrinal analysis with structured artefact evaluation. Legal and regulatory analysis of the European Union, the United States, and the World Intellectual Property Organization framework is used to derive governance requirements, while existing provenance and traceability systems are reviewed to establish the technical landscape and inform the design. On that basis, the research develops a workflow-level provenance approach, including a new Entity-Action-Attribution data model and an extensible metadata structure for representing governance-relevant information such as AI involvement, rights reservations, and authorisation conditions. The approach is implemented and demonstrated in representative application scenarios, including a multimodal chatbot-based environment, and evaluated against the requirements derived from the legal and technical analysis.

Keywords: provenance, Human-AI-created works, attribution, copyright, AI disclosure, data governance, traceability, interoperability

The originality of this thesis has been checked using the Turnitin Originality Check service.

Use of AI in thesis

I have utilised AI tools in my thesis:

☐ No

☒ Yes

The AI tools utilised in my thesis and their purposes are described below:

Names and versions of AI tools: Anthropic Claude Code (Opus 4.6); ChatGPT (Plus Version 5.4); Grammarly

Purpose of using AI tools: Anthropic Claude Code was used as a coding assistant within Visual Studio Code during the implementation of the prototype and related technical components. It was used to support programming tasks such as code drafting, refactoring, debugging, and clarifying implementation options. ChatGPT was used mainly to improve mechanical aspects of writing, including wording, sentence structure, clarity, flow, and organisation of draft text. It was also used to support revision of sections by helping refine phrasing and presentation. Grammarly was used for proofreading and improving grammar, spelling, punctuation, and sentence-level clarity.

Sections where AI tools were used: AI tools were used mainly for coding assistance in prototype development and for language revision, sentence-level improvement, and editorial refinement across the document.

I acknowledge that I am fully responsible for the entire content of my thesis, including the parts generated by AI, and accept accountability for any violations of ethical standards in publications.

Table of Contents

Abstract ii

AI Use Disclosure iii

Table of Contents iv

List of Figures vii

List of Tables viii

List of Abbreviations ix

Chapter 1: Introduction 1

1.1 Background 1

1.2 Research Problem 2

1.3 Research Aim 2

1.4 Research Questions 2

1.5 Significance of the Study 3

1.6 Scope and Limitations 3

Chapter 2: Literature Review 4

2.1 Intellectual Property, Authorship, and Human-AI Creation 4

2.1.1 Intellectual Property and Generative AI 4

2.1.2 Authorship and Copyrightability in Human-AI Works 4

2.1.3 Copyright Infringement, Licensing, and Responsibility as Information Problems 7

2.2 AI Disclosure and Traceability 9

2.3 Provenance as an Approach for Process Documentation 10

2.3.1 Defining Provenance 10

2.3.2 Taxonomic Dimensions of Provenance Systems and Design Considerations 10

2.3.3 Provenance Documentation in Explainable and Trustworthy AI 12

2.4 Existing Technical Approaches to AI Content Identification and Traceability 13

2.4.1 AI Detection Tools 13

2.4.2 Watermarking and Embedded AI-Origin Signals 13

2.4.3 Content Authenticity and Asset-Centric Traceability 14

2.5 W3C PROV as an Interoperability Model for Provenance Representation 15

2.6 Privacy, Integrity, and Selective Disclosure in Provenance Systems 15

2.7 Research Gap 16

Chapter 3: Research Methodology 18
3.1 Research Design 18
3.1.1 Problem Identification 19
3.1.2 Definition of Objectives 19
3.1.3 Design and Development 20
3.1.4 Demonstration 20
3.1.5 Evaluation 20
3.1.6 Communication 22
3.2 Methods 22
3.2.1 Normative Doctrinal Analysis 22
3.2.2 Structured Artefact Evaluation 23
3.2.3 Data Collection 23
3.2.4 Analysis 24
3.3 Limitations 24
Chapter 4: Results 25
4.1 Governance Requirements Derived from the Legal and Technical Analysis 25
4.2 Provenance Information Structures Required for Human-AI Governance 28
4.2.1 The EAA Data Model 29
4.3 Design of the Proposed Provenance Approach 30
4.3.1 The Extensibility System 31
4.3.2 The Privacy Module 32
4.3.3 The Smart Contract Layer 33
4.3.4 The Storage and Indexer Layer 34
4.3.5 The API and Integration Interface 34
4.4 Demonstration 35
4.4.1 Inline Provenance Disclosure 36
4.4.2 Session-Level Provenance Recording 37
4.4.3 The Graph-Based Derivation View 38
4.4.4 Resource-Level Recording 38
4.4.5 Action-Level Recording 39
4.4.6 Typed Participant Representation 40
4.4.7 Attribution Representation 41
4.5 Evaluation 41
4.5.1 Evaluation Against R1 to R6 42
4.5.2 Structural Evaluation of the Approach 42
4.5.3 Supporting Implementation Evidence 42
4.5.4 Limitations of the Artefact 43
4.6 Summary 43

Chapter 5: Discussion 44
5.1 Alignment with Legal, Regulatory, and Technical Literature 44
5.2 Design Choices and Their Rationale 45
5.3 Extent of Support and Its Limits 46
5.4 Intrinsic Limitations of Provenance as a Governance Mechanism 47
5.5 Contributions 48
Chapter 6: Conclusion 49
6.1 Answers to the Research Questions 49
6.2 Significance and Practical Implications 49
6.3 Limitations and Constraints 50
6.4 Future Works 50
References 51
Appendices 56
Appendix A1: Gantt Chart 56
Appendix A2: Test Suite Summary 57

List of Figures

Figure 4.1. Architecture of the proposed provenance approach 31

Figure 4.2. Content Provenance popover triggered from the inline Pr badge 36

Figure 4.3. Session Provenance panel open alongside an active conversation.....37

Figure 4.4. Full provenance graph in the Provenance 38

Figure 4.5. Resources view of the provenance record.....39

Figure 4.6. Actions view of the provenance record.....40

Figure 4.7. Entities view of the provenance record 40

Figure 4.8. Attribution view of the provenance record.....41

List of Tables

Table 4.1. Governance requirements, regulatory basis, and primary approach components ..27

List of Abbreviations

AI – Artificial Intelligence

API – Application Programming Interface

C2PA – Coalition for Content Provenance and Authenticity

CJEU – Court of Justice of the European Union

CID – Content Identifier

DID – Decentralised Identifier

DSRM – Design Science Research Methodology

EAA – Entity-Action-Attribution

ECDSA – Elliptic Curve Digital Signature Algorithm

EU – European Union

EVM – Ethereum Virtual Machine

GDPR – General Data Protection Regulation (EU) 2016/679

IPFS – InterPlanetary File System

IP – Intellectual Property

MPC – Multi-Party Computation

OPM – Open Provenance Model

REST – Representational State Transfer

SDK – Software Development Kit

TEE – Trusted Execution Environment

TDM – Text and Data Mining

US – United States

W3C – World Wide Web Consortium

WIPO – World Intellectual Property Organization

Chapter 1: Introduction

1.1 Background

Digital content is produced and distributed through complex technical and social systems and plays a central role in modern communication, commerce, and cultural production. Increasingly, such content is created through workflows that involve multiple contributors, tools, and platforms rather than a single act of authorship. Text, images, audio, video, and code are often produced through sequences of drafting, generation, editing, transformation, and redistribution. In many contexts, understanding how a work was produced is therefore important for attribution, accountability, and governance.

This issue has become more significant with the growth of generative artificial intelligence. Generative AI is now used within iterative creative workflows in which human users prompt systems, review outputs, select among alternatives, and refine results over multiple steps. Human-AI creation may include prompting, reviewing outputs, editing generated material, and combining AI-generated and human-authored elements across several stages or platforms. As a result, the final artefact may reveal little about the process through which it came into being.

Recent legal and regulatory developments reflect the growing importance of this process information. Copyright analysis of AI-assisted works depends on the nature and extent of human contribution (U.S. Copyright Office, 2025). The EU AI Act requires certain AI-generated or AI-manipulated content to be identifiable in machine-readable form (Regulation [EU] 2024/1689), while the DSM Directive requires rights reservations against certain forms of text and data mining to be expressed in machine-readable form if they are to operate effectively at scale (Directive [EU] 2019/790, art. 4). Across these frameworks, governance depends on information about inputs, transformations, actors, and authorisation conditions.

Data provenance offers one possible response to this need. In digital systems, provenance refers to the recorded history of data or artefacts, including their origins, transformations, dependencies, and associated actors (Pérez et al., 2018). Although provenance has been studied extensively in areas such as scientific workflows, databases, and linked data, its application to Human-AI content creation remains underdeveloped. This thesis therefore examines how a provenance approach can be designed to support attribution, disclosure, licensing, and responsibility in Human-AI creation workflows.

1.2 Research Problem

Transparency in how content is created is becoming increasingly important in digital environments, especially in the age of generative artificial intelligence. Questions of authorship, disclosure, licensing, and responsibility cannot be answered adequately in the absence of reliable process-level information. There is a growing expectation in legal and regulatory frameworks that Human-AI creation processes should be made legible in machine-readable form, such that relevant actors can be identified, AI involvement can be disclosed and traced, rights reservations can be recognised, and the production history of a work can be reconstructed when needed.

Existing technical approaches address only parts of this problem. AI detection tools infer likely AI origin from outputs. Watermarking methods signal AI involvement at the point of generation. Content authenticity systems support asset-level traceability, while general provenance models provide interoperable representation structures. No single approach, however, fully integrates workflow-level capture, governance-relevant metadata, interoperability, and privacy-aware disclosure for Human-AI creations.

1.3 Research Aim

The aim of this study is to design and evaluate a governance-oriented provenance approach for Human-AI-created-works that can support attribution, credit assignment, and responsibility in digital content creation workflows.

1.4 Research Questions

This thesis addresses the following research questions:

RQ1. What legal and governance requirements shape attribution, responsibility, liability, and transparency in relation to Human-AI-created works?

RQ2. What types of provenance information are required to capture and document Human-AI content creation workflows in a way that supports attribution and accountability?

RQ3. How can an interoperable and privacy-aware provenance approach be designed to capture the entities, actions, artefacts, derivations, and governance-relevant conditions involved in Human-AI creation processes?

RQ4. To what extent can the proposed provenance approach support attribution, disclosure, licensing compliance, and responsibility assessment in representative Human-AI creation scenarios?

1.5 Significance of the Study

As AI systems become embedded in widely used tools and platforms, understanding how digital artefacts are produced becomes important for governance, licensing, and compliance. Regulatory frameworks are beginning to require greater transparency regarding the creation and use of AI-generated content. For example, recent regulatory developments introduce obligations related to the identification and traceability of AI-generated outputs (Regulation [EU] 2024/1689). Meeting these expectations requires technical mechanisms capable of capturing information about the creation process.

Provenance systems provide one such mechanism by recording the origin and transformation of digital artefacts. While provenance has been studied extensively in areas such as data management and scientific workflows, its application to Human-AI creative processes remains an emerging area of research. By designing and evaluating a provenance approach tailored to Human-AI content creation workflows, this study contributes to ongoing efforts to improve transparency and accountability in AI-mediated content production.

1.6 Scope and Limitations

This thesis focuses on the design and evaluation of a provenance approach for Human-AI-created works. It is concerned with recording and structuring process-level information about content creation workflows, not with explaining the internal reasoning of generative AI models.

The legal and regulatory analysis is limited to selected frameworks that are especially influential for this problem: the European Union, the United States, and the international intellectual property context represented by the World Intellectual Property Organization. This scope is justified by the transnational character of digital content creation and AI deployment, but it does not amount to a comprehensive comparative survey of all jurisdictions.

The approach is demonstrated through representative scenarios and prototype integrations rather than through large-scale real-world deployment. As a result, the thesis establishes design feasibility and conceptual adequacy, but does not claim to establish long-term operational performance, adversarial robustness, or full legal sufficiency in regulatory or litigation settings.

Chapter 2: Literature Review

This chapter reviews the literature that informs the design of a provenance approach for Human-AI-created works. It first examines the legal and regulatory context governing authorship, infringement, licensing, disclosure, and accountability in generative AI. It then reviews technical work on provenance, traceability, and attribution systems, focusing on how existing approaches model creative workflows, record inputs and outputs, and support interoperability and privacy preservation.

2.1 Intellectual Property, Authorship, and Human-AI Creation

This section reviews the legal concepts most relevant to Human-AI-created works. It considers the intellectual property context, the question of authorship and copyrightability, and the ways in which infringement, licensing, and responsibility become problems of process-level information.

2.1.1 Intellectual Property and Generative AI

Intellectual property refers to creations of the mind, such as inventions, literary and artistic works, designs, symbols, names, and images used in commerce. It is commonly divided into industrial property and copyright. Industrial property includes patents for inventions, trademarks, industrial designs, and geographical indications, while copyright covers literary works, films, music, artistic works, architectural designs, and computer programs (World Intellectual Property Organization [WIPO], 2020). Intellectual property rights allow creators or owners of patents, trademarks, and copyrighted works to benefit from their own work or investment in a creation.

The advent of generative artificial intelligence poses a fundamental challenge to existing intellectual property frameworks and to their enforcement. Generative AI systems can produce outputs that resemble what copyright law has traditionally treated as human creative expression. This raises difficult questions about authorship, ownership, and the allocation of rights.

2.1.2 Authorship and Copyrightability in Human-AI Works

Traditional legal understandings of authorship have long been anchored in the assumption that creative works arise from human agency, with intellectual property doctrines justifying

ownership and exclusive rights in human originality and intent (Omogiate, 2024). The law has long accommodated the use of tools and machines in creative production, such as a camera or notation software, without attributing authorship to the instrument itself. Generative AI, however, does not merely assist or record the expression of a human author. It produces outputs by selecting and structuring expressive elements in ways that go beyond the historically passive role of creative tools (U.S. Copyright Office, 2025).

The United States Copyright Act of 1976 provides that "copyright in a work protected under this title vests initially in the author or the authors of the work" (17 U.S.C. § 201(a)). Although the statute does not expressly define the term "author", statutory interpretation and longstanding doctrine have consistently treated human creative agency as a necessary condition for copyright protection. The Act's subject-matter provision also defines protected works as those that may be perceived "either directly or with the aid of a machine or device" (17 U.S.C. § 102(a)).

The proposition that AI systems cannot themselves be recognised as legal authors has been tested directly in recent case law. In Thaler v. Perlmutter, the U.S. Copyright Office refused registration for an artwork identified as having been generated solely by an AI system, and that refusal was upheld on judicial review (Thaler v. Perlmutter, 2023, 2025). The courts held that copyright protection under the Copyright Act requires human authorship. This line of authority is important not only because it rejects AI authorship, but also because it makes the identification of human contribution legally significant.

European Union law reflects a similar orientation through the doctrine of originality. According to the Court of Justice of the European Union (CJEU), a work is protected only if it is the "author's own intellectual creation," meaning that it reflects the author's free and creative choices and expresses the author's personality (Infopaq International A/S v Danske Dagblades Forening, C-5/08, 2009; Painer v Standard VerlagsGmbH, C-145/10, 2011; Football Dataco Ltd v Yahoo! UK Ltd, C-604/10, 2012). An AI system that generates content through statistical analysis of training data does not make choices in this human sense, nor does it express a personality. For that reason, it cannot meet the EU originality standard. Although EU copyright law does not expressly address AI authorship, its emphasis on human creativity strongly suggests that only humans can qualify as authors.

At the international level, neither the Berne Convention nor WIPO practice requires a different conclusion. WIPO discussions on intellectual property and AI have reflected broad recognition that extending copyright protection to fully AI-generated works would require deliberate policy change rather than ordinary interpretation of existing law (WIPO, 2020).

The rejection of AI systems as legal authors does not mean that works created with AI assistance are necessarily excluded from copyright protection. Rather, the legal question turns on whether the human user contributed original expression or exercised sufficient creative control over the resulting work. The U.S. Copyright Office's (2025) report on copyrightability and artificial intelligence addresses this issue in detail. Issued after a formal Notice of Inquiry and informed by more than 10,000 public comments, the report reflects a substantial consultative process and has become an important institutional account of the current U.S. position on AI-assisted copyrightability. Its central position is that copyright may subsist in the human-authored aspects of a work, even where AI-generated material is present, provided that the human contribution satisfies the required threshold of originality (U.S. Copyright Office, 2025).

This is clearest where a human contributes a protectable expression directly. If a user inputs their own copyrightable material into an AI system and that material remains perceptible in the final output, copyright may continue to subsist in those human-created elements (U.S. Copyright Office, 2025). Protection may also arise where the human makes original choices in selecting, coordinating, arranging, or modifying AI-generated material. By contrast, mechanical, trivial, or purely technical edits are not sufficient. Not every intervention amounts to authorship. Whether an AI-assisted work satisfies the originality threshold, therefore, depends on the specific character and extent of the human contribution in each case (U.S. Copyright Office, 2025). Under the current understanding of generative systems, purely AI-generated material, or material in which the human role consists only of giving general instructions through simple prompts without controlling the expressive result, falls outside copyright protection.

The literature therefore suggests that authorship and copyrightability in Human-AI works are process-dependent. What matters is not only the finished artefact, but also the process through which it came into being. The work's production history becomes legally significant. This includes prompts, intermediate outputs, acts of selection or rejection, edits, the incorporation

of pre-existing material, and the sequence of transformations through which the final artefact emerged. The literature on authorship thus points to the need for records capable of distinguishing human and AI contributions across identifiable artefacts and stages of the creative workflow.

2.1.3 Copyright Infringement, Licensing, and Responsibility as Information Problems

The literature on infringement, lawful use, licensing, and responsibility is broader and more unsettled than the literature on authorship, especially where debates concern training data and outputs generated by AI models.

Copyright law gives the owner the economic right to authorise or prevent certain uses of a work. The owner of a work can authorise or prohibit its reproduction in all forms, public performance, recording, broadcasting, translation, and adaptation (WIPO, n.d.). Unauthorised use of this nature constitutes copyright infringement. Generative AI has intensified these debates because such systems are typically trained on vast amounts of data that may include copyrighted works, such as text, images, and other media (Andersen v. Stability AI Ltd, 2023; Getty Images (US), Inc. v. Stability AI, Ltd., 2023). In addition, generative AI systems may produce outputs that resemble copyrighted works contained within the training dataset, thereby creating possible infringement concerns during training and after deployment.

In the European Union, AI training may engage the reproduction right under Directive 2001/29/EC and, in some circumstances, the sui generis database right under Directive 96/9/EC. The DSM Directive introduced text and data mining exceptions that are especially relevant here. Article 3 establishes a mandatory exception for research organisations and cultural heritage institutions, while Article 4 extends a broader exception to other users, including commercial actors, subject to an opt-out mechanism through which right holders may reserve their rights (Directive [EU] 2019/790, arts. 3–4). The practical importance of Article 4 lies in the fact that rights reservations must be communicated in machine-readable form if they are to be recognised and acted upon at scale.

The EU AI Act reinforces this compliance logic. Article 53(1)(c) requires providers of general-purpose AI models to implement policies ensuring compliance with EU copyright law, including respect for rights reservations made under Article 4(3) of the DSM Directive (Regulation [EU] 2024/1689, art. 53[1][c]). Article 53(1)(d) further requires the publication of

a sufficiently detailed summary of the content used for training, including copyright-protected material (Regulation [EU] 2024/1689, art. 53[1][d]). Recital 107 makes clear that this summary should be detailed enough to enable right holders and the public to assess relevant copyright implications, while still taking account of trade secret concerns. These provisions do not resolve all legal questions surrounding training-data use, but they indicate that compliance and accountability depend on structured information about what material was used and on what basis.

In the United States, AI training has largely been analysed through the fair use framework in 17 U.S.C. § 107, and the U.S. Copyright Office (2025) observes that the legality of such uses remains unsettled and context-dependent. Even so, the literature makes clear that assessing licensing and compliance depends on knowing what materials were used, whether permission existed, whether rights were reserved, and how those materials entered the model or workflow.

Questions about whether an AI-generated output reproduces protected expression, whether it derives from earlier works in a legally relevant sense, or whether it substitutes for licensable uses of protected material often require more than inspection of the final output alone (U.S. Copyright Office, 2025). Assessment may depend on lineage, intermediate transformations, source inputs, or the presence of pre-existing human-authored materials in the workflow. The legal answer in any given case will depend on jurisdiction and facts, but the literature suggests that such assessment becomes significantly more difficult where the creation process is opaque.

A similar pattern appears in discussions of content involving identifiable persons, voices, or likenesses. Copyright traditionally protects literary, artistic, and audiovisual works that are fixed in a tangible medium of expression. While generative AI exposes a distinct category of harm in the form of unauthorised replication of a person's voice, image, or likeness, copyright law is not designed to address this. Copyright does not protect an individual's identity as such, and courts have made clear that a person's name, likeness, or unfixed voice does not automatically fall within copyright subject matter (Downing v. Abercrombie & Fitch, 2001; Midler v. Ford Motor Co., 1988).

Where content creation involves identifiable persons, voices, likenesses, or consent-sensitive materials, questions may arise regarding the scope of authorisation, duration, and permitted use. Recent discussions of digital replicas illustrate that certain uses depend not only on copyright status, but also on consent and authorisation conditions (U.S. Copyright Office,

2024). Whether these issues are framed through data protection law, contract, platform governance, or proposed digital replica legislation, they indicate the importance of artefact-level records of permission and use conditions.

Questions of responsibility and liability bring these informational issues into even sharper focus. Where disputes arise over harmful, misleading, infringing, or unauthorised content, responsibility may depend on differentiating the roles played by users, platform operators, model providers, and other actors within the workflow (U.S. Copyright Office, 2024, 2025). Liability questions, therefore, overlap with traceability questions: who initiated the process, who supplied inputs, which system generated which output, what later modifications were made, and under what conditions those acts occurred. Although the literature does not converge on a single liability model for all Human-AI contexts, it consistently suggests that responsibility cannot be assessed meaningfully where the relevant chain of actions and actors is not recorded.

2.2 AI Disclosure and Traceability

The need for process-related information is also visible in recent regulations concerning AI disclosure and traceability, especially in the transparency provisions of the EU AI Act.

Article 50(2) of Regulation (EU) 2024/1689 requires providers of AI systems capable of generating synthetic text, image, audio, or video content to ensure that outputs are marked in a machine-readable format and are detectable as artificially generated or manipulated, insofar as this is technically feasible (Regulation [EU] 2024/1689, art. 50[2]). The provision further states that the technical solutions used for this purpose should be effective, interoperable, robust, and reliable. Recital 133 identifies metadata, watermarks, cryptographic methods, logging methods, fingerprints, and provenance-related techniques as possible means of satisfying this obligation. The obligation does not apply in every case. It excludes certain assistive editing functions and certain law-enforcement uses, but for systems producing substantial synthetic content, it remains a meaningful and enforceable requirement.

The AI Act's provisions on general-purpose AI models reinforce the same point. Article 53(1)(d) requires the publication of a sufficiently detailed summary of training data content, and Recital 107 indicates that this disclosure should enable right holders and the public to assess relevant copyright implications while balancing transparency against legitimate trade secret protection (Regulation [EU] 2024/1689, art. 53[1][d]; recital 107). Taken together, these

provisions introduce disclosure requirements at both the model level and the output level. They do not, by themselves, create complete process-level provenance. They do, however, reflect a regulatory recognition that governance in the AI context depends on information that is structured, persistent, and capable of supporting later scrutiny.

2.3 Provenance as an Approach for Process Documentation

This section turns from legal and regulatory demands to provenance as a technical response to those demands. It defines provenance, reviews major provenance design considerations, and considers its relevance to explainable and trustworthy AI.

2.3.1 Defining Provenance

In its earliest use, provenance referred to the documented history of ownership of a work of art. It described the chain of custody through which a painting, sculpture, or artefact could be traced back to its creator, thereby supporting claims of authenticity and influencing market value. The term has retained related meanings in other contexts. In the food industry, provenance refers to the traceable geographic and agricultural origin of a product. In archival practice, it is a foundational principle of arrangement, requiring records to be preserved in relation to the person, body, or organisation that created them.

Within information technology and computer science, provenance has been formalised as a concept applicable to digital data. Pérez et al. (2018) define provenance as the total body of information consisting of all elements and relationships that contribute to the existence of a piece of data. The technical literature commonly distinguishes between two basic forms of provenance. Source provenance concerns the data used to create a data item, that is, the inputs and sources from which it was derived. Transformation provenance concerns the processes applied to those inputs, namely the sequence of operations through which the data reached its present form.

2.3.2 Taxonomic Dimensions of Provenance Systems and Design Considerations

Pérez et al. (2018), in a systematic review of 105 studies covering 25 representative provenance systems, propose a six-dimensional taxonomy that has informed later work in the field. The dimensions are general aspects, data capture, data access, subject, storage, and non-functional requirements. These dimensions are useful both for evaluating existing systems and for identifying what a provenance system for Human-AI created works must be capable of doing.

The first dimension, general aspects, describes the overall background of provenance systems. Pérez et al. explain that this dimension includes categories such as provenance definition, data-processing category, application domain, purpose, whether a proposal is an intended provenance system or an extension of an existing tool, and software availability. Under the data-processing category, they distinguish workflow-management-system-based, script-based, service-based, and query-based approaches. The emphasis here is on the general character of a provenance system: the context in which it operates, the kind of data processing it supports, and the purpose for which it was developed, such as error tracing, quality estimation, audit trail, or understanding derived data.

The data capture dimension concerns how provenance information is collected. Within this dimension, the central distinction is between eager and lazy provenance collection. Eager approaches compute and store provenance information as the relevant process occurs. This creates runtime overhead, but produces complete records that are immediately available. Lazy approaches defer provenance computation until it is requested. This reduces storage and runtime costs during execution, but may restrict the forms of provenance that can later be reconstructed. Most workflow-based systems, including VisTrails, Taverna, and Kepler, use eager capture. Database-oriented systems such as ZOOM and Cui 2000 use lazy approaches. Capture may also occur at the workflow, activity, operating system, service, or database level, each of which involves different trade-offs between granularity, completeness, and dependence on particular technologies (Pérez et al., 2018).

The data access dimension concerns how users access provenance repositories for examination. Pérez et al. distinguish between browsing and API-based access, and between exploratory and directed querying. In exploratory access, users browse provenance information without a fully specified question in advance, while in directed access they formulate specific provenance questions using a query language. The review also notes that provenance systems may use general languages such as SQL, XQuery, XPath, or SPARQL, or may provide dedicated provenance-query mechanisms.

Within the subject dimension, the taxonomy distinguishes among different forms of provenance according to content, including lineage, where-provenance, why-provenance, and how-provenance, and according to granularity, orientation, and interoperability. Fine-grained provenance records metadata at the level of individual data elements. Coarse-grained

provenance records derive histories at the workflow level. Process-oriented models focus on derivation processes and on their inputs and outputs. Data-oriented models focus more directly on data dependencies. The interoperability category identifies OPM and PROV as the main approaches for exchanging and understanding provenance records across different systems.

The storage dimension concerns how provenance information is registered, stored, and maintained over time. Pérez et al. describe this dimension in terms of storage scalability, coupling, persistence, and archiving. They distinguish between centralised and distributed storage approaches, and between high-coupling, loose-coupling, and no-coupling strategies depending on the relationship between provenance data and the data whose provenance is being recorded. They also note that provenance systems use a variety of persistence models, with relational databases, XML, and RDF being especially common. In relation to archiving, the review distinguishes between sequences of deltas and time-stamping as approaches to representing changes over time. Provenance storage therefore concerns not only where provenance is kept, but also how it is linked to the underlying data and how change is recorded across time.

The non-functional requirements dimension covers security/privacy, verification, and repeatability/reproducibility/replayability. Pérez et al. note that security is central to the trustworthiness of provenance data and identify confidentiality, integrity, unforgeability, non-repudiation, and availability as key concerns. They also distinguish privacy from security by treating it as the controlled management and limitation of sensitive provenance information.

2.3.3 Provenance Documentation in Explainable and Trustworthy AI

Kale et al. (2023) extend the relevance of provenance into the literature on explainable and trustworthy AI. Their review argues that provenance documentation can support transparency and trust in AI-based systems by preserving information about data origins, processing steps, workflow structure, and responsible actors. In their account, provenance answers the who, what, when, and where of AI and machine-learning workflows and helps users trace how a result came into being.

A central point in Kale et al.'s analysis is that post-hoc explanations alone are not sufficient for explainable and trustworthy AI. Explanation methods may describe model behaviour or provide local reasons for a prediction, but they do not necessarily preserve the wider workflow

context through which that result emerged. Provenance contributes something different. It records process history, metadata, and relationships among artefacts, activities, and actors, making it possible to examine how data and results were produced. Kale et al. therefore present provenance as complementing rather than replacing explainability techniques.

Kale et al. also connect provenance to repeatability, reproducibility, comparability, and trustworthiness. They note that repeatability and comparability in AI and machine-learning experiments depend on understanding metadata and provenance, and they argue that provenance can contribute to fairness, accountability, transparency, and explainability in AI systems. Their discussion is especially useful here because it shows why provenance matters not only in scientific data management but also in AI settings where workflow opacity creates problems of trust and interpretation.

2.4 Existing Technical Approaches to AI Content Identification and Traceability

Not all technical responses to AI-content governance are provenance systems in the strict sense. Some methods analyse final outputs and infer whether they are likely to be AI-generated. Others embed origin signals into the output. Others attach signed assertions to the asset itself. These approaches differ in what they capture and in the kind of evidence they provide.

2.4.1 AI Detection Tools

AI detection tools operate at the level of the final output. They analyse a text, image, audio clip, or video and estimate whether it is likely to have been generated by an AI system. Such tools may be useful in screening, moderation, or triage contexts where direct provenance information is unavailable. Their limitation is that they are retrospective and probabilistic. They infer likely origin from the properties of the artefact rather than recording the process through which it was produced (Weber-Wulff et al., 2023).

For that reason, AI detection is not equivalent to provenance. A detector may suggest that an output appears machine-generated, but it cannot directly show which model produced it, what prompt was used, what intermediate versions existed, or what human intervention followed.

2.4.2 Watermarking and Embedded AI-Origin Signals

A second class of approaches seeks to identify AI-generated content by embedding a signal into the output at or near the point of generation. These methods are commonly described as

watermarking or AI-origin signalling. Because they rely on generation-time embedding rather than later inference alone, they can support stronger forms of detectability than output-only classifiers.

Google DeepMind describes SynthID as a tool “to watermark and identify content generated through AI” and states that it is designed to help users identify AI-generated or altered content, thereby fostering transparency and trust in generative AI. The same source explains that SynthID embeds digital watermarks directly into AI-generated images, audio, text, or video, and that these watermarks are intended to be imperceptible to humans while remaining detectable through corresponding identification technology. For images and video, Google DeepMind states that the watermark is added at the moment content is created and is designed to remain detectable after common modifications such as cropping, filters, frame-rate changes, or lossy compression (Google DeepMind, n.d.).

Watermarking contributes directly to AI disclosure and detectability. Even so, its scope remains limited. A watermark may indicate that AI was involved in producing an artefact, but it does not normally record prompt history, alternative outputs, user edits, incorporated source materials, or the rights and authorisation conditions associated with those elements. Watermarking is therefore better understood as an output-level disclosure mechanism than as a full process-level provenance solution (Google DeepMind, n.d.).

2.4.3 Content Authenticity and Asset-Centric Traceability

A third class of approaches provides richer information than watermarking by attaching signed assertions to the digital asset itself. The most important example is the work of the Coalition for Content Provenance and Authenticity (C2PA). The C2PA technical specification defines a C2PA Manifest as the set of information about the provenance of an asset based on assertions, a claim, and a claim signature. It also defines provenance as the logical concept of understanding the history of an asset and its interaction with actors and other assets. In addition, the specification states that provenance can disclose how an asset was created, how it was changed, and what was changed, and that when an asset changes, its existing provenance is preserved while new changes are added (C2PA, 2026).

This makes C2PA especially relevant because it does support edit history and change tracking. It provides a stronger model of traceability than either AI detection or watermarking alone by

combining asset-level assertions with tamper-evident linkage between the asset and its associated metadata. At the same time, C2PA remains primarily asset-centric. Its core unit is the digital asset and its manifest, rather than the full workflow across prompts, alternative generations, user selections, actor-role distinctions, or broader governance metadata such as rights reservations and consent scope. In this respect, C2PA captures an important part of the problem, but not the entire process history that Human-AI governance may require (C2PA, 2026).

2.5 W3C PROV as an Interoperability Model for Provenance Representation

Within the broader provenance literature, the W3C PROV family is especially important because it provides a recognised model for representing and exchanging provenance information. The W3C defines provenance as information about entities, activities, and people involved in producing a piece of data or thing, and states that the goal of PROV is to enable the interoperable publication and interchange of provenance information in heterogeneous environments such as the Web (World Wide Web Consortium [W3C], 2013a).

Kale et al. (2023) also discuss PROV as a key model for provenance documentation in AI-related contexts. They describe the PROV Data Model as including entities and activities with timing information, derivation relations, responsible agents, bundles supporting provenance of provenance, and collections providing logical structure. Their review also notes the existence of formats and tools that support exchange, validation, and visualisation of PROV records.

The relevance of PROV here is not that it determines the final approach in advance, but that it offers a widely recognised interoperability model. Its concepts are broad enough to describe many of the structural relations that matter in Human-AI workflows. PROV is intentionally generic. It does not, by itself, specify how to model AI disclosure metadata, rights reservations, consent scope, contribution weighting, or selective disclosure of sensitive workflow information (W3C, 2013a).

2.6 Privacy, Integrity, and Selective Disclosure in Provenance Systems

The provenance literature also shows that recording process history is not enough on its own. Provenance records must be trustworthy and usable under conditions of privacy and confidentiality. Pérez et al. (2018) note that provenance is of limited value for audit evidence if security concerns are not addressed, and they identify confidentiality, integrity,

unforgeability, non-repudiation, and availability as important security properties for provenance data. They also distinguish privacy from security by treating it as the controlled management and limitation of sensitive provenance information. At the same time, they report that security, privacy, and verification remain comparatively weakly supported across the systems they reviewed, with many implementations relying on the host environment and none of the surveyed systems using formal verification methods.

Kale et al. (2023) connect this issue to trustworthy and explainable AI more broadly. Their review argues that provenance documentation contributes to transparency, traceability, explainability, and reproducibility in AI-based systems, but also suggests that provenance must remain usable under different stakeholder needs and contexts. In that sense, provenance must support trust without assuming unrestricted exposure of all process data.

2.7 Research Gap

The literature establishes, with reasonable convergence, that governance of Human-AI content creation is fundamentally an information problem. Copyright doctrine in both the United States and the European Union has made the legal status of AI-assisted works turn on the character and extent of human contribution across identifiable stages of the creative process (U.S. Copyright Office, 2025; Infopaq, 2009; Painer, 2011). Regulatory frameworks, most notably the EU AI Act (Regulation [EU] 2024/1689) and the DSM Directive (Directive [EU] 2019/790), reinforce this demand by requiring that AI involvement be disclosed in machine-readable form and that training-data use be documented sufficiently for rights holders to assess copyright implications. Questions of licensing compliance, output-level infringement, and responsibility attribution depend in turn on knowing what materials entered a workflow, what transformations they underwent, and who authorised what at each stage. Across these governance contexts, the ability to answer legal and regulatory questions rests not on the finished artefact alone but on a reliable record of the process through which it came into being.

Against this established need, the technical literature reveals approaches that are individually valuable but collectively insufficient. AI detection tools operate retrospectively and probabilistically, inferring likely AI origin from output properties rather than recording the process that produced them. Watermarking, as exemplified by SynthID (Google DeepMind, n.d.), contributes meaningfully to disclosure and detectability at the point of generation, but it does not capture prompt history, intermediate outputs, user selections, or the rights and

authorisation conditions associated with incorporated source materials. The C2PA specification (2026) advances matters further by attaching tamper-evident assertions to digital assets and supporting edit history at the artefact level, yet it remains asset-centric: its core unit is the digital asset and its manifest rather than the full workflow across actors, tools, generative invocations, and governance-relevant metadata. The W3C PROV model (W3C, 2013a) offers the most credible foundation for interoperable provenance exchange but is intentionally domain-agnostic and does not specify how to represent the entities and relationships particular to Human-AI creation workflows, nor how to model rights reservation status, consent scope, or the selective disclosure requirements that privacy and confidentiality constraints impose.

The gap is therefore not simply the absence of a provenance tool. Several tools exist, and each addresses a genuine part of the problem. The gap is the absence of a unified, governance-oriented approach that integrates what those tools address only partially and in isolation. The literature does not yield a approach that coherently specifies workflow-level provenance requirements across the distinct governance objectives of authorship assessment, infringement evaluation, licensing compliance, regulatory disclosure, and responsibility attribution. Nor does it provide one that maps those requirements onto a structured and interoperable data model, treats rights reservation status and authorisation conditions as first-order provenance metadata, and addresses how governance-relevant verification can be made available to authorised parties without unrestricted exposure of commercially sensitive or personally identifiable workflow data.

Pérez et al. (2018) observe that security, privacy, and verification remain the most weakly supported properties across existing provenance systems, and Kale et al. (2023) argue that provenance must remain usable across different stakeholder needs and contexts. Taken together, the legal, regulatory, and technical literature points to the same conclusion: the informational infrastructure that Human-AI content governance presupposes has not yet been designed in a form that is simultaneously complete, trustworthy, and usable in the decentralised, multi-stakeholder environment in which that governance must operate.

Chapter 3: Research Methodology

This chapter presents the research methodology adopted for the design, development, and evaluation of a provenance approach intended to support attribution, credit assignment, responsibility, and liability for Human-AI-created works. It outlines the conceptual foundations of the research, the overall design, and the methods used to translate legal and technical requirements into a technical artefact.

3.1 Research Design

This study adopts Design Science Research Methodology, hereafter DSRM, as its overarching research design. DSRM is appropriate where research aims to address an identified problem through the purposeful design and evaluation of an artefact, while also contributing knowledge about the problem and the solution space (Peffers et al., 2007–2008). In the present study, the artefact is a governance-oriented provenance approach for Human-AI-created works, developed in response to a governance gap at the intersection of copyright, AI disclosure, and data governance.

The choice of DSRM follows from the nature of the research problem. The thesis does not only seek to describe legal and technical challenges surrounding Human-AI-created works. But also seeks to respond to those challenges by designing and evaluating an approach that can support attribution, credit assignment, responsibility, and governance in digital content creation workflows. This makes DSRM especially suitable, since the methodology is intended for research that moves from problem identification to artefact creation and assessment. It is also well suited to socio-technical problems in which the artefact must respond to both normative and technical requirements and in which the contribution is expected to be both practical and scholarly.

Peffers et al. (2007–2008) describe DSRM as comprising six related stages: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication. In the present study, they provide the overall structure through which legal and technical findings are translated into a provenance approach and through which that approach is later examined. DSRM therefore functions as a framework for linking problem analysis, artefact design, operational implementation, and evaluation within a coherent research process.

3.1.1 Problem Identification

The first DSRM stage involved defining and motivating the problem addressed in the thesis. The problem was identified as the absence of a governance-oriented provenance approach capable of supporting attribution, responsibility, liability, disclosure, and related governance functions in Human-AI-created works. Existing legal and regulatory frameworks increasingly require answers to questions about how a work was produced, who contributed to it, what materials entered the workflow, and under what conditions relevant acts occurred. At the same time, existing technical systems tend either to focus on outputs rather than workflows or to provide provenance structures without the governance-relevant metadata needed in Human-AI contexts.

The motivation for the study therefore arises from a growing mismatch between governance demands and available informational infrastructure. Effective allocation of credit or accountability requires a structured understanding of the creation process itself. Where that process is opaque, governance becomes difficult, fragmented, or speculative. This stage of the research was informed by review and analysis of copyright doctrine, AI disclosure obligations, data governance frameworks, and technical literature on provenance, traceability, and content authenticity.

3.1.2 Definition of Objectives

The second DSRM stage involved defining the objectives that a viable solution would need to satisfy. These objectives were derived from the legal and technical analysis and later formalised in Chapter 4 as six functional requirements, R1 to R6. At a general level, the proposed solution had to satisfy five broad objectives. First, it had to support sufficient workflow granularity to distinguish relevant acts, contributors, and artefacts. Second, it had to support interoperable representation so that provenance information could function across systems and contexts. Third, it had to support governance-relevant metadata such as AI disclosure, rights reservations, and authorisation conditions. Fourth, it had to address privacy and controlled disclosure, since governance may require proof of certain provenance facts without unrestricted exposure of the full workflow. Fifth, it had to be capable of operational implementation rather than remaining only a conceptual model. These objectives translated the research problem into criteria against which the artefact could later be designed and evaluated.

3.1.3 Design and Development

The third DSRM activity is design and development. Peffers et al. describe this stage as the creation of the artefact itself, which may take the form of constructs, models, methods, or instantiations. They also note that the artefact should embed a research contribution in its design. In this study, the design and development stage was guided by the defined objectives and focused on the creation of a research artefact comprising a conceptual provenance approach and an accompanying technical implementation. This stage involved designing a conceptual provenance model capable of representing entities, actions, inputs, outputs, and their relationships within Human-AI creation workflows.At the design stage, the system architecture, data flows, and storage structures were specified, with particular attention to interoperability and privacy. On that basis, an operational implementation was developed to realise the proposed approach and to support systematic evaluation of its functionality and applicability.

3.1.4 Demonstration

The fourth DSRM activity is demonstration. Peffers et al. define demonstration as showing the use of the artefact to solve one or more instances of the identified problem. They note that this may take the form of experimentation, simulation, case study, proof, or another appropriate activity. The operational implementation of the developed provenance system was applied and demonstrates through a chatbot-based AI assistant application with multimodal capabilities. This environment was selected to reflect contemporary and widely adopted modes of interaction with generative AI systems. The demonstration application also included functionality enabling users to construct workflows involving multiple models. The demonstration was designed to show the system's ability to capture provenance information across multiple sessions and multi-step workflows, and to trace inputs, outputs, actions, and associated entities in an environment representative of current generative AI usage patterns.

3.1.5 Evaluation

The fifth DSRM activity is evaluation. Peffers et al. describe evaluation as observing and measuring how well the artefact supports a solution to the problem. They also stress that evaluation involves comparing the objectives of the solution with the actual results from use

of the artefact, and that the form of evaluation may vary depending on the artefact and research setting.

The evaluation process assessed the extent to which the proposed provenance system satisfied the defined objectives of granularity, interoperability, privacy preservation, and support for attribution and responsibility assignment in Human-AI created works. First, the evaluation criteria were specified in advance. The six governance requirements, R1 to R6, were derived from the normative doctrinal analysis and fixed prior to implementation. A requirement was considered satisfied only where the approach provided a named and functional structural component that addressed the corresponding governance deficit without relying on ad hoc or implementation-specific workarounds. This pre-specification is important because it ensures that the evaluation is conducted against externally grounded requirements rather than criteria defined retrospectively to fit the artefact.

Second, the evaluation was conducted under controlled and representative conditions. The artefact was examined through a chat-based demonstration environment designed to reflect a contemporary Human-AI workflow while remaining sufficiently bounded for systematic assessment. This made it possible to evaluate whether the approach could record typed human and AI participation, represent actions, artefacts, and derivation relations across workflow steps, attach governance-relevant metadata, and support privacy-aware disclosure and verification in a working prototype setting. The purpose of this stage was to assess design adequacy and functional support, not to claim large-scale operational validation.

Third, the evaluation was supported by implementation-level test evidence. In addition to the requirements-based and scenario-based assessment, the operational implementation was accompanied by an automated test suite that provided objective and reproducible evidence of software quality. At the time of submission, the suite comprised 1,099 passing tests, including smart contract tests, storage tests, indexer tests, extension tests, payment adapter tests, privacy module tests, git-provenance tests, media and C2PA tests, SDK tests, and API integration tests. These tests support claims of technical operability and specification conformance. Their role in the evaluation is therefore evidentiary and supportive: they strengthen the reliability of the implementation assessed, while the primary basis of evaluation remains the pre-specified governance requirements and their operational demonstration. A package-level summary of the

test suite is provided in Appendix A2, and the tests are reproducible by running “pnpm test” from the repository root.

3.1.6 Communication

The sixth DSRM activity is communication. Peffers et al. describe this activity as communicating the problem, the artefact, its utility and novelty, the rigour of its design, and its effectiveness to relevant audiences. The outcomes of this research, including the design of the provenance approach, its operational implementation, and the results of its demonstration and evaluation, are documented in this thesis. Selected implementation materials have also been made publicly accessible through the ProvenanceKit website, code repository, developer documentation, and the deployed demonstration application. These materials provide access to the implemented artefact and support transparency and further engagement.

3.2 Methods

While DSRM provides the overarching research structure, this section describes the specific methods used to operationalise the methodology. Two complementary methods were employed: normative doctrinal analysis and structured artefact evaluation.

3.2.1 Normative Doctrinal Analysis

Normative doctrinal analysis was employed to examine the legal frameworks governing intellectual property and data governance as they relate to authorship, responsibility, and liability in the context of creative output and artificial intelligence. The analysis was primarily grounded in legislative and regulatory instruments and in the interpretation of statutory provisions.

The analysis was conducted across three jurisdictions selected because, collectively, they represent and inform prevailing international approaches to intellectual property, data governance, and artificial intelligence: the European Union, the United States, and the international treaty framework administered by the World Intellectual Property Organization. This scope was necessary given the inherently cross-border character of intellectual property and the transnational infrastructure within which artificial intelligence systems are developed, deployed, and used.

Within these jurisdictions, legal material was examined to determine the extent to which existing frameworks address four analytically distinct but interrelated concerns: the attribution of authorship to Human-AI outputs; the allocation of legal responsibility and liability for such outputs; the governance of personal and training data implicated in the creation process; and the conditions under which provenance records could serve as legally meaningful evidence.

The purpose of this normative doctrinal analysis was to extract legally grounded requirements across intellectual property and data governance domains that a provenance system must satisfy in order to support attribution, credit assignment, responsibility determination, and lawful disclosure.

3.2.2 Structured Artefact Evaluation

Structured artefact evaluation was applied to identify and analyse design patterns and interoperability characteristics in existing systems exhibiting provenance-related or attribution-related functionality. This method was oriented toward a specific design question: what structural and technical properties must the proposed provenance system possess in order to integrate meaningfully with the systems through which Human-AI creation currently takes place?

Systems were selected on the basis of three criteria: public accessibility through standard interfaces or documented application programming interfaces (APIs); the presence of at least one provenance-related or attribution-related feature evidenced by peer-reviewed literature or official documentation; and representativeness of tool categories implicated in contemporary generative AI workflows. Examination of these systems was instrumental in grounding design decisions in the observable realities of the current technical landscape.The evaluation was conducted through direct interaction with publicly accessible interfaces and inspection of publicly available technical documentation. No system-internal data was accessed, and no user data was collected or analysed.

3.2.3 Data Collection

Legal materials were obtained from authoritative and publicly accessible legal databases and institutional repositories. European Union directives and regulations were accessed through EUR-Lex, which provides official texts of EU legislation. International intellectual property treaties, policy documents, and guidance materials were sourced from the WIPO-Lex

repository. United States federal case law and judicial materials were obtained from official resources provided by the United States Court.

Secondary technical and scholarly literature was collected from established academic databases and indexing services. Peer-reviewed research articles and conference proceedings were sourced from the IEEE Xplore digital library, while additional literature was identified using Google Scholar. These sources were used to identify relevant work on data provenance, Human-AI collaboration, privacy-preserving systems, and emerging technical approaches.

Observations derived from direct examination of deployed systems, treated as primary research artefacts, constituted original data. No datasets were extracted from the examined systems, and no internal system data or user data were accessed at any stage.

3.2.4 Analysis

Findings from the normative doctrinal analysis and the structured artefact evaluation were qualitatively analysed independently and subsequently synthesised to inform the design and evaluation of the proposed provenance system as presented in chapter 4.

3.3 Limitations

Several limitations apply to the research methodology and methods employed in this study. First, the normative doctrinal analysis focuses on selected legal frameworks and jurisdictions and does not provide an exhaustive comparative analysis of all global intellectual property or data governance regimes. While the selected jurisdictions are representative of influential legal approaches, the findings may not fully reflect jurisdiction-specific variations beyond the scope of this study.

Second, the structured artefact evaluation is limited to observable system behaviour and publicly available documentation. The analysis does not assess internal system architectures, proprietary mechanisms, or undisclosed implementation details, which may affect the completeness of conclusions regarding certain technical capabilities.

Third, the evaluation of the proposed provenance system is conducted through scenario-based demonstration rather than large-scale real-world deployment. As a result, conclusions regarding scalability, long-term operational performance, and adoption dynamics are necessarily limited.

Chapter 4: Results

This chapter presents the results of applying Design Science Research Methodology (DSRM) to the problem of provenance for Human-AI-created works. The central contribution is a governance-oriented provenance approach whose requirements, data model, and verification mechanisms are derived from legal and regulatory obligations rather than from generic data-management concerns. The literature review established that no existing system meets these criteria collectively. The results reported here address that gap across six requirement dimensions.

4.1 Governance Requirements Derived from the Legal and Technical Analysis

The first result of the study is the derivation of a concrete set of governance requirements from the broader legal and technical problem identified in earlier chapters. This directly addresses the first research question (RQ1) by identifying the legal and governance requirements that shape attribution, responsibility, liability, and transparency in relation to Human-AI-created works. The literature review established that governance of Human-AI-created works depends on more than the existence of a final artefact. Across authorship doctrine, copyright compliance, AI disclosure obligations, and questions of liability and consent, the decisive issue is whether the underlying process can be reconstructed in a structured and reliable manner.

Five governance objectives were identified from the analysis: authorship assessment, infringement evaluation, licensing compliance, regulatory disclosure, and responsibility attribution. Each of these objectives reveals a corresponding informational deficit in current Human-AI creation environments.

In line with the objective-definition stage of DSRM, these governance deficits were translated into six functional requirements, designated R1 to R6. These requirements define the minimum capabilities that the artefact must possess in order to address the identified problem in an operationally meaningful way. They serve as the bridge between problem identification and artefact design by converting the broader governance challenge into a set of explicit design criteria. As such, they provide the basis for the approach architecture developed in the following sections and for the later demonstration and evaluation of the artefact against its intended purpose.

The six requirements are as follows.

R1 requires the per-operation recording of human and AI acts as typed, distinct events linked to identified content artefacts. This requirement follows from the need to distinguish human contribution from AI generation at the artefact level.

R2 requires machine-readable licensing metadata, including an AI training opt-out field queryable by content identifier, together with the recording of authorisation conditions as first-order provenance metadata. This requirement follows from the need to represent rights reservations and permissions in a form that can operate across digital systems.

R3 requires complete, traversable derivation chains linking transformed outputs to their upstream inputs. This requirement follows from the need to trace transformations, dependencies, and potential reuse of prior works across the workflow.

R4 requires structured AI metadata that distinguishes AI-as-tool from AI-as-agent contexts and remains machine-readable and interoperable across systems. This requirement follows from the need to disclose not only that AI was involved, but also the form and character of that involvement.

R5 requires per-artefact authorisation metadata recording consent status, grantor, scope, and expiry. This requirement follows from the need to document consent-sensitive uses, especially where works involve identifiable persons or other restricted content.

R6 requires selective, cryptographically verifiable disclosure of provenance claims without requiring full record exposure. This requirement follows from the need to reconcile governance and accountability with data minimisation, confidentiality, and restricted access.

Taken together, these requirements establish the minimum governance-oriented capabilities that the proposed provenance approach must provide. They therefore form the basis for the provenance information structures developed in the next section and for the architectural design presented later in the chapter.

Table 4.1 maps the identified governance deficits to the six functional requirements (R1–R6) and indicates the primary approach component through which each requirement is addressed.

Table 4.1. Governance requirements, regulatory basis, and primary approach components

Regulatory Basis	Informational Deficit	Requirement	Primary Component
Copyright authorship doctrine (US: 17 U.S.C. § 102; EU: Infopaq C-5/08; U.S. Copyright Office, 2025)	No per-operation record distinguishes human contribution from AI generation at the artefact level	R1: Per-operation recording of human and AI acts as typed, distinct events linked to identified content artefacts; entity roles differentiate human actors from AI systems	Typed workflow records in the EAA model, supported by provenance event anchoring
DSM Directive Art. 4(3); EU AI Act Art. 53(1)(c)(d)	Rights reservations against AI training use are not machine-readable or queryable by content identifier; authorisation conditions are undocumented at the artefact level	R2: Machine-readable licensing metadata, including an AI training opt-out field queryable by content identifier; authorisation conditions recorded as first-order provenance metadata	Licensing and authorisation extensions attached to provenance records. ext:license@1.0.0; ext:authorization@1.0.0
Copyright liability (17 U.S.C. § 107; Thaler v. Perlmutter, 2025)	When AI systems transform or aggregate prior works, the derivation chain is not recorded or recoverable	R3: Complete, traversable derivation chains linking transformed outputs to their upstream inputs	Action-based derivation modelling and graph-based provenance traversal
EU AI Act Art. 50(2), Recital 133	AI-generated outputs are not marked in machine-readable, interoperable form; the degree and nature of AI involvement are unrepresented	R4: Structured AI metadata distinguishing AI-as-tool from AI-as-agent; machine-readable; cross-platform interoperable	AI-specific extension metadata within an interoperable provenance bundle. ext:ai@1.0.0 (dual tool/agent modes; autonomy level);
GDPR Art. 9; EU AI Act Art. 50(3); proposed digital replica legislation	Consent for content depicting real persons is not documented at the artefact level	R5: Per-artefact authorisation metadata recording consent status, grantor, scope, and expiry	Per-artefact authorisation metadata within the extension layer ext:authorization@1.0.0
GDPR Arts. 5(1)(c), 25 (data minimisation); commercial confidentiality	Full exposure of provenance records conflicts with data minimisation and confidentiality obligations	R6: Selective, cryptographically verifiable disclosure of provenance claims without requiring full record exposure	Privacy module: selective disclosure; Pedersen commitments; @noble/ciphers (Cure53-audited)

Three additional design considerations emerged from the technical literature and shaped the approach alongside R1 to R6. These were interoperability with established provenance concepts, support for heterogeneous storage infrastructures, and stronger treatment of privacy and verification than is commonly provided in existing systems. These considerations did not replace the six governance requirements, but informed how they were implemented.

4.2 Provenance Information Structures Required for Human-AI Governance

The second result of the study concerns the types of provenance information required to support governance once the requirements have been established. It substantively addresses the second research question (RQ2) by showing what kinds of information must be captured and documented in Human-AI content creation workflows to support attribution and accountability. The findings show that generic provenance recording is not sufficient on its own. A governance-oriented approach must combine workflow-level provenance structures with metadata that carries legal and institutional significance.

Five categories of provenance information were found to be necessary. First, the approach must represent participants in the creation process, including human actors, AI systems, and organisations. Governance depends on distinguishing these roles rather than treating them as undifferentiated agents. Second, it must represent content artefacts across all stages of the workflow, including prompts, generated outputs, intermediate versions, and transformed derivatives. Governance questions frequently concern not only the final output, but also the materials and versions from which it emerged. Third, it must represent actions through which artefacts are produced, modified, selected, aggregated, or transformed. The findings show that governance requires granularity at the level of individual operations rather than only coarse histories of origin. Fourth, it must represent derivation relationships linking inputs, transformations, and outputs across workflow stages. This is necessary for tracing influence, dependency, and transformation across the lifecycle of a work. Fifth, it must represent governance-relevant metadata, including attribution, AI involvement, rights reservations, consent and authorisation conditions, and contribution weighting. These are not peripheral annotations. They are part of what makes the provenance record useful for governance.

4.2.1 The EAA Data Model

The provenance information requirements identified in the previous section are realised in the **Entity Action Attribution (EAA)** model, which forms the conceptual core of the proposed approach. The EAA model is designed to represent Human-AI creation workflows in a way that supports attribution, accountability, and governance. It does so by specialising general provenance concepts for the Human-AI creative domain. Whereas W3C PROV provides a general vocabulary of entities, activities, and agents, the EAA model introduces typed participant roles, first-class attribution structures, and extension points for governance-specific metadata. In this way, it provides a provenance model better suited to the specific informational demands identified in RQ2**.**

An **Entity** represents a participant in the production process. Its role field distinguishes human, ai, and organization actors. This typed distinction is the primary mechanism through which human and AI participation is recorded as structurally different rather than inferred retrospectively from context.

A **Resource** represents a content artefact of any media type. Each resource is addressed through a stable content reference, with a content-addressed identifier used as the preferred scheme. This provides the stable per-item identity needed for rights queryability, derivation tracking, and cross-platform reference.

An **Action** is the central unit of workflow granularity. Each generation, transformation, aggregation, selection, or other operation produces a distinct action record linking typed inputs to typed outputs through the performing entity. This supports process reconstruction at the level required for attribution and accountability.

An **Attribution** is modelled as a first-class object rather than merely a property attached to a final artefact. It links an entity to an action or resource in a named role and can carry further metadata, such as contribution weights and authorisation context.

A **Provenance Bundle** assembles the model components into a signed and exchangeable container with a shared context URI, enabling records to be interpreted independently of the implementation that produced them.

4.3 Design of the Proposed Provenance Approach

The governance requirements identified and the provenance information structures developed establish what the proposed approach must be capable of representing. The third result of the study concerns the architectural design of an operational provenance approach. At this stage, the study addresses the third research question (RQ3) by showing how the identified requirements and information structures can be realised operationally to capture entities, actions, artefacts, derivations, and governance-relevant conditions in Human-AI workflows.

The proposed approach is organised as a layered architecture centered on the EAA model and extended through distinct but coordinated components for semantic extensibility, controlled disclosure, anchoring and verification, storage and indexing, and application-level integration. The layered design preserves a stable core provenance data structure while enabling the approach to remain extensible, privacy-aware, independently verifiable, and deployable across heterogeneous Human-AI environments.

This architectural organisation reflects five design principles. First, the core provenance model must remain stable and implementation-independent so that records can be interpreted across different systems and contexts. Second, governance-specific metadata must be attachable without altering the underlying model, since different applications require different provenance semantics. Third, privacy and accountability must be addressed together, so that relevant provenance claims can be verified without requiring unrestricted disclosure of the full workflow. Fourth, trust in the provenance record should not depend solely on a single platform operator where later scrutiny or contestation may arise. Fifth, the approach must remain deployable across heterogeneous infrastructures and usable within existing Human-AI systems. The remainder of this presents the principal architectural components through which these design principles are realised. Figure 4.1 shows the layered architecture showing the application layer, API and integration interface, core EAA provenance model, extensibility system, privacy module, storage and indexer layer, smart contract layer, and the governance outcomes supported by the approach

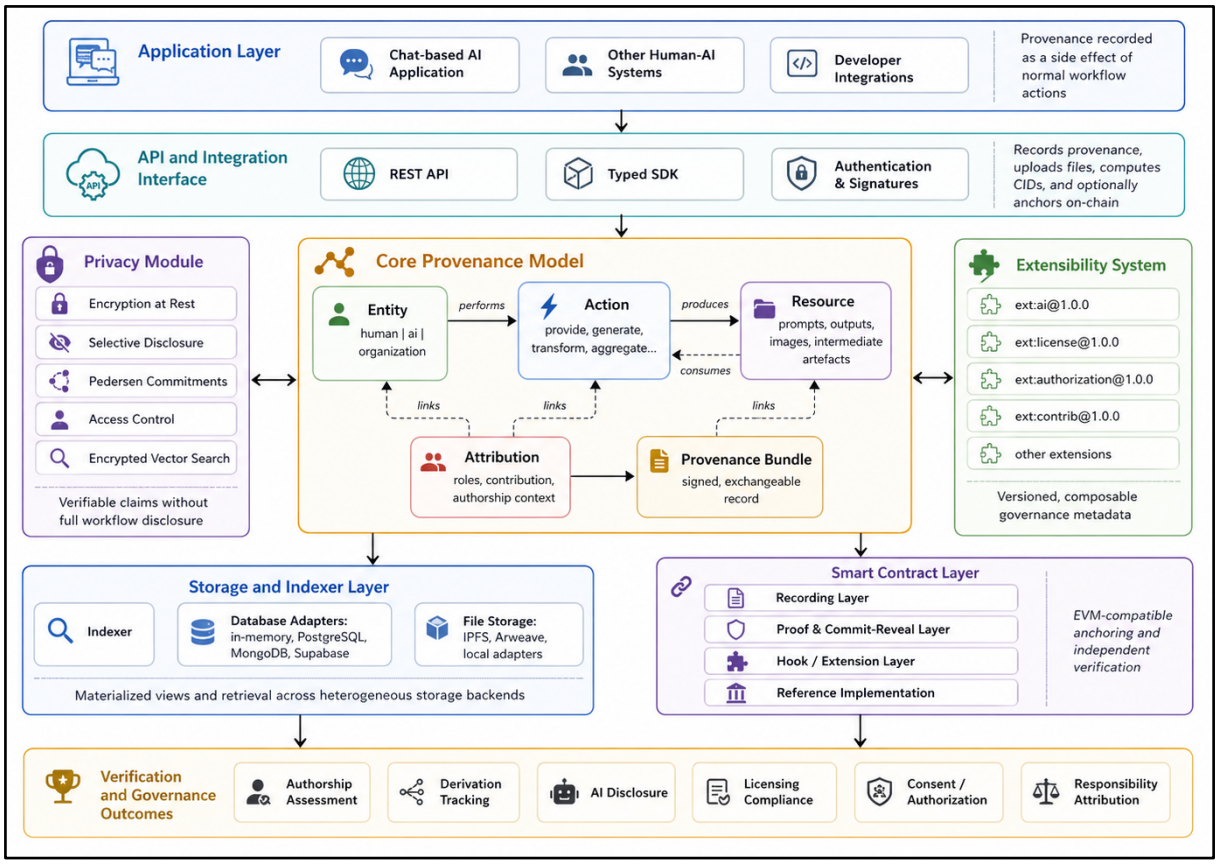


Figure 4.1. Architecture of the proposed provenance approach

4.3.1 The Extensibility System

The EAA base model is intentionally minimal. It defines the structure of provenance records without embedding domain-specific governance semantics. Domain content is added through a typed extension system in which structured metadata is attached to any EAA record under a versioned namespace key of the form `ext:namespace@semver`. Extensions are additive and composable.

This design operationalises the requirement identified by Kale et al. (2023) that provenance infrastructure must remain usable across different stakeholder needs and contexts. A platform deploying the approach for AI disclosure purposes may attach `ext:ai@1.0.0`, while a rights-management platform may attach `ext:license@1.0.0`; both consume the same interoperable base record.

The governance-critical extensions are as follows.

The `ext:ai@1.0.0` satisfies R1 and R4 through a dual-mode design. In tool mode, attached to an Action, it records provider, model, prompt fingerprint, generation parameters, and token counts. In agent mode, attached to an AI Entity, it records the AI system's autonomy level, delegation context, and session identifier. The two modes are structurally necessary because they require different metadata payloads: tool mode describes a discrete invocation, while agent mode describes the persistent configuration of an autonomous participant. A single-mode structure would require fields to be silently ignored depending on context.

The `ext:license@1.0.0` satisfies R2. Its `aiTraining` field (`permitted` | `reserved` | `unspecified`) renders a DSM Article 4(3) opt-out machine-readable and queryable by content identifier, without requiring natural-language licence terms to be parsed.

The `ext:authorization@1.0.0` satisfies R2 and R5 by recording authorisation status, grantor, scope, expiry, and cryptographic proof per artefact, thereby providing the consent documentation infrastructure required by GDPR Article 9, EU AI Act Article 50(3), and proposed digital replica legislation.

The `ext:contrib@1.0.0` records relative contribution weights on Attribution objects, normalised to basis points using the Largest Remainder Method. This supports both authorship analysis and revenue allocation from the same provenance data.

The remaining extensions cover blockchain anchoring, proof attestation, identity verification, server-side witness attestation with optional Trusted Execution Environment reports, tool-level attestation, per-claim verification status, ownership lifecycle management, payment routing, content pinning, and micropayment access control.

4.3.2 The Privacy Module

A second architectural requirement is that provenance must support accountability without requiring unrestricted disclosure of the full workflow. Because Human-AI provenance may contain personal, commercially sensitive, or otherwise restricted information, the approach must allow specific provenance claims to be disclosed, verified, or contested without exposing the entire record. The privacy module addresses this requirement by enabling selective disclosure under verifiable conditions. In this way, privacy is treated as an architectural property of the approach rather than as an external constraint on it. The privacy and selective disclosure requirement R6 is realised through five mechanisms.

Encryption at rest uses the `@noble/ciphers` library (Cure53-audited), with XChaCha20-Poly1305 as the default algorithm. Resources are encrypted before storage; the CID remains a public provenance anchor while the content itself is key-gated.

Selective disclosure allows specific provenance claims, such as creation date, licence status, or attributed entity, to be proven cryptographically without exposing the full bundle. This implements GDPR data minimisation at the provenance layer.

Pedersen commitments over the `secp256k1` curve enable verification of numeric claims, such as contribution weights, without revealing the committed values. Their homomorphic property allows a set of weighted attributions to be verified as summing correctly without disclosing individual weights.

Access control supports token-gated provenance access through a provider interface, enabling institutional and licensed-content use cases across multiple EVM-compatible networks.

Encrypted vector search encrypts embedding vectors client-side before storage, enabling semantic search over provenance records without exposing prompt or content embeddings to the storage server.

4.3.3 The Smart Contract Layer

A third architectural requirement is that trust in the provenance record should not depend exclusively on the assertions of a single operator. Because provenance in this context may later be used to support attribution disputes, compliance checks, or questions of responsibility, the approach requires a means of independently verifying that certain acts were recorded and when they were recorded. The smart contract layer addresses this requirement by anchoring provenance events in a public, permissionless verification substrate where their existence and ordering can be checked without relying on the recording operator alone.

The approach implements a four-layer smart contract hierarchy deployable on any EVM-compatible blockchain, with each layer adding capabilities without altering the semantics defined beneath it. The first layer is a base recording layer that emits immutable, indexed events for every production act. The second is a proof and commitment layer that supports ECDSA signature verification and a hash-based commit-reveal scheme for cases in which commercial sensitivity requires deferred disclosure. The third is an extensible hook layer through which

third-party contracts can receive callbacks on every recorded action without modifying the core contracts. The fourth is a full reference implementation that adds entity registration, resource registration, attribution recording, and permissive ownership management.

The commit-reveal scheme on the proof layer is architecturally distinct from the Pedersen commitment mechanism in the privacy module. The former is a hash-based, on-chain construct for deferred disclosure, while the latter is an off-chain cryptographic scheme for verifying numeric claims without revealing their values.

Two chain adapters, one built on the `viem` library and one implementing the EIP-1193 standard directly, allow the approach to be used with any compatible wallet provider, including browser-extension wallets, embedded wallets, and multi-party computation wallets, without imposing a specific client-library dependency.

4.3.4 The Storage and Indexer Layer

Storage adapters are materialised views of the authoritative blockchain event log rather than primary records. Any adapter can be independently repopulated from the public event log. This preserves trustworthiness under adversarial conditions: a contested or corrupted database does not compromise the underlying provenance record.

The approach defines pluggable storage interfaces for both provenance data and file storage. Four database backends are provided, namely in-memory, PostgreSQL, MongoDB, and Supabase, with the Supabase adapter additionally supporting vector similarity search and client-side encrypted vector storage for semantic provenance queries. Six file-storage adapters cover the principal IPFS pinning services, Arweave, and local IPFS. A dedicated indexer component bridges the on-chain event log and the off-chain storage layer, performing historical synchronisation from the deployment block and real-time polling thereafter. It handles five event types through pure transformation functions, producing EAA-typed records that are then stored via the pluggable storage interface.

4.3.5 The API and Integration Interface

The approach exposes a REST API over two authenticated namespaces: an end-user namespace for provenance recording and retrieval, and a management namespace for operator-level administration. The separation ensures that end-user provenance traffic and platform

administration operate under independent authentication and permission controls, supporting the multi-stakeholder model identified in the literature review.

A typed client SDK wraps the API, both chain adapters, and signing utilities. Its primary method performs file upload, CID computation, off-chain provenance recording, and optional non-blocking on-chain anchoring in a single call, reducing the integration burden for application developers to a small number of function calls at the application layer.

4.4 Demonstration

The artefact was demonstrated through a multi-provider conversational AI prototype built on the proposed provenance approach. The application was developed by the author and is presented as a demonstration of functional deployment rather than as independent validation. It provides a representative Human-AI interaction setting in which users can submit prompts, receive text responses, generate images, and inspect the resulting provenance record through an integrated provenance interface.

The purpose of the demonstration was to show how the proposed approach can be operationalised within an ordinary content-creation environment in what is the current typical ai application interaction mode without requiring users to leave the application or manually construct provenance records. In this implementation, provenance recording occurs as a side effect of normal interaction. When a user submits a prompt, the system records a provide action linking the human entity to the prompt resource. When the model returns a response, the system records a generate action linking the AI entity to the output resource. For example, where an image is generated from a text prompt, an additional derivation relation is recorded between the textual input and the image output. At the same time, governance-relevant metadata, including AI model information, prompt fingerprints, and token counts, is attached through the extension system. In this way, the demonstration operationalises the core elements of the proposed approach: typed entities, workflow-level actions, content-addressed resources, derivation tracking, attribution structures, and extension-based metadata.

The demonstration also shows how this information can be surfaced to users in a form that supports both immediate transparency and deeper inspection. At the interface level, the application exposes two related provenance views. First, an inline provenance badge allows users to open a compact disclosure view for a specific output. Second, a dedicated provenance

explorer provides access to the full session-level record, including resources, actions, entities, attributions, and the chronological session timeline. Together, these views show how the same provenance record can support both lightweight disclosure and fuller workflow reconstruction.

4.4.1 Inline Provenance Disclosure

The first interface-level manifestation of the provenance approach appears directly within the conversation through an inline provenance badge rendered on AI-generated outputs. Selecting this badge opens the **Content Provenance** popover shown in **Figure 4.2**. The popover presents structured disclosure metadata for the selected output, including the creation date, the producing entity, and the AI tool used.

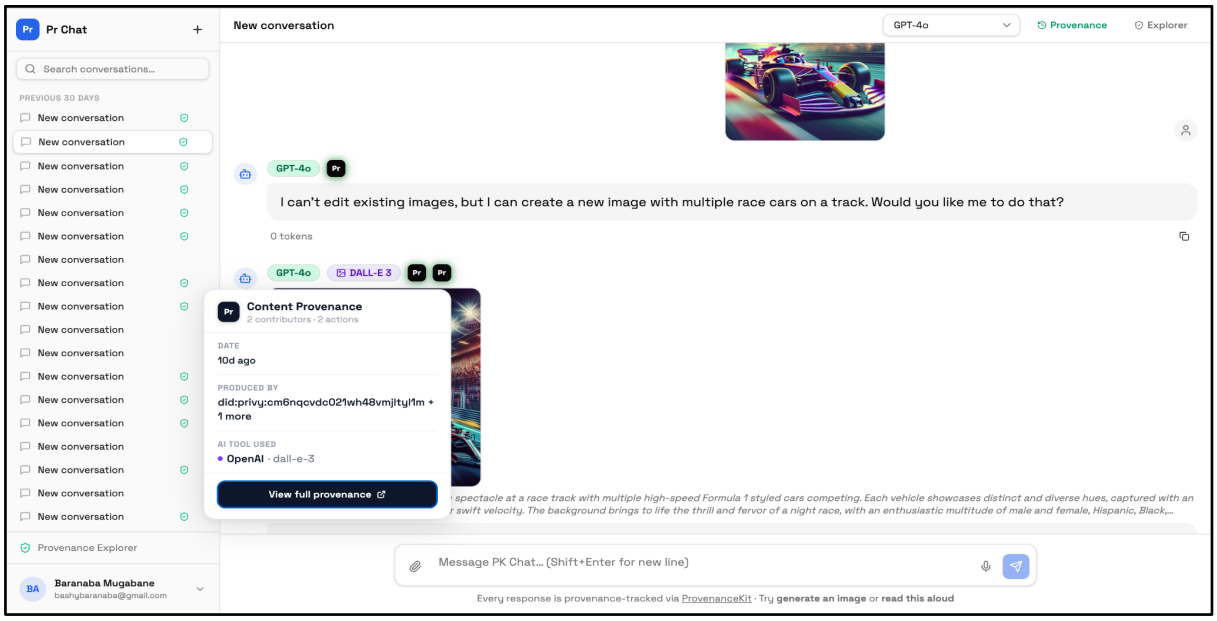


Figure 4.2. Content Provenance popover triggered from the inline Pr badge

This view demonstrates how the proposed approach can support machine-readable AI disclosure in a user-facing form. Rather than requiring a separate compliance interface or external credential viewer, the disclosure information is generated directly from the provenance record attached to the output. In practical terms, this means that the same provenance record that supports later workflow reconstruction can also support immediate disclosure at the point where the content is encountered. This directly illustrates prototype-level support for R4, while also showing how the provenance approach can be embedded within a normal conversational workflow.

4.4.2 Session-Level Provenance Recording

Beyond per-output disclosure, the application records provenance at the level of the interaction session as a whole. Rather than treating each output as an isolated event, it records the conversation as an evolving workflow in which prompts, responses, revisions, selections, and further generations are linked together over time. This is important because Human-AI creation often involves trying out different prompts, reviewing alternatives, refining earlier results, and building new outputs from prior steps. In this implementation, those stages are recorded in-session as they occur, preserving not only the final outputs but also the iterative development through which they emerged. The Session Timeline view in Figure 4.3 renders this session-level provenance as a chronological, turn-by-turn record of the conversation.



Figure 4.3. Session Provenance panel open alongside an active conversation

This view is important because it makes visible how the EAA model operates in practice. A human entity performs a provide action by submitting a prompt resource. An AI entity then performs a generate action that produces a text output resource. Where image generation occurs, a further generate action by a tool-specific AI model produces an image resource linked back to the prompt through a derivation relation. Each step is therefore not only stored, but also presented in sequence, showing how one action leads to the next. The timeline view demonstrates that the approach records provenance at workflow level rather than only at the level of final outputs. It therefore shows how process reconstruction becomes possible in a normal Human-AI interaction setting and provides prototype-level support for R1 and R3.

4.4.3 The Graph-Based Derivation View

The dedicated Provenance Explorer also surfaces the full provenance graph for any recorded resource. Whereas the Session Timeline view presents the workflow chronologically, the Graph view presents it structurally by showing how entities, actions, and resources are linked within the provenance record. Figure 4.4 shows the graph for a generated image, tracing the derivation chain across the session.

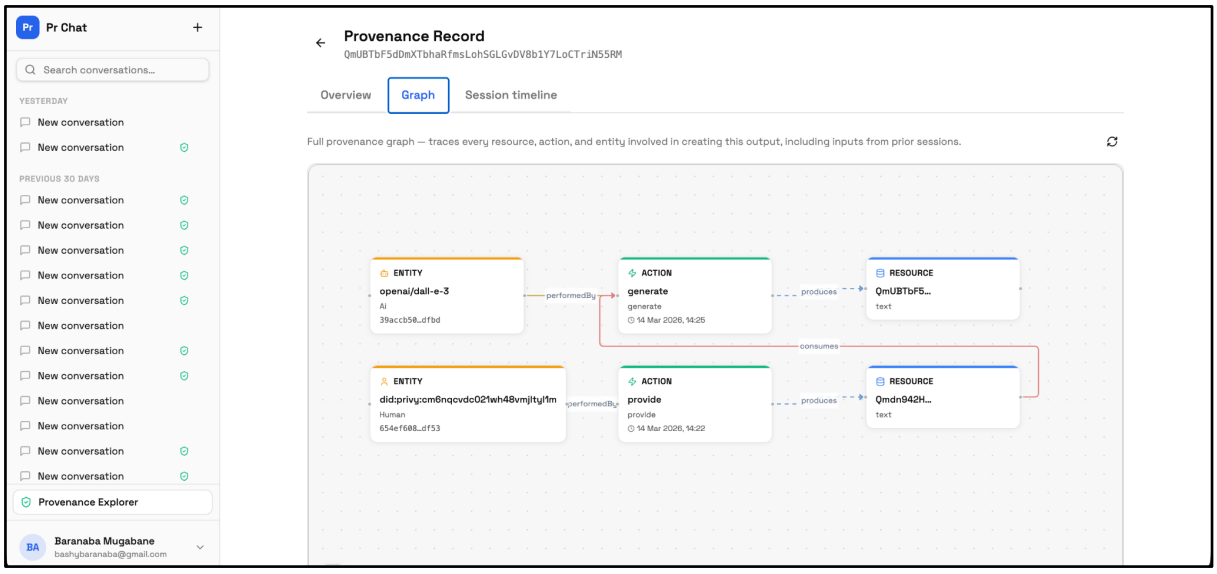


Figure 4.4. Full provenance graph in the Provenance

In this scenario, a human entity performs a provide action that produces the textual prompt resource, while an AI entity performs a generate action that consumes that prompt and produces the final image resource. The graph therefore makes explicit that the output is the result of a recorded chain of actions and dependencies rather than an isolated artefact. This directly demonstrates traversable derivation, typed participant distinction, and stable content-addressed identity, thereby providing prototype-level support for R1 and R3.

4.4.4 Resource-Level Recording

The Resources view surfaces the stored artefacts associated with the provenance bundle. These include text and image resources created over the course of the conversation and recorded as content-addressed items within the provenance storage layer. As shown in Figure 4.5, the visible entries include both TEXT and IMAGE resources, each referenced through a truncated content identifier and linked to storage.

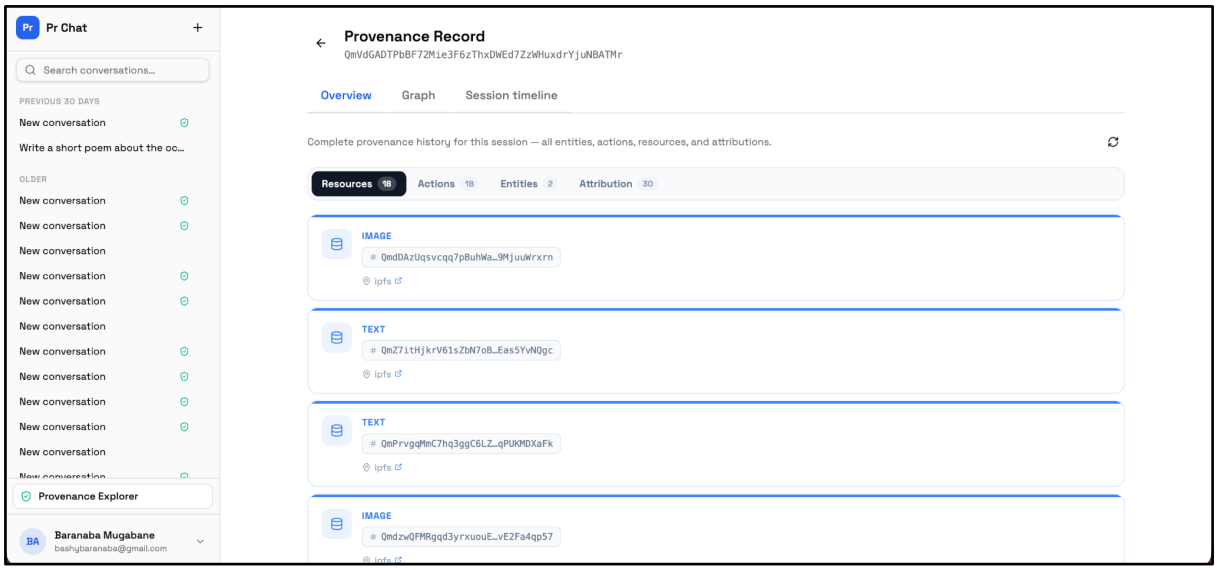


Figure 4.5. Resources view of the provenance record

This view demonstrates that the system records not only actions and participants, but also the concrete artefacts through which the workflow can later be reconstructed. In relation to the EAA model, it provides interface-level evidence of the Resource layer. In relation to the governance requirements, it supports stable artefact identity, derivation tracing, and later ability to query specific content items.

4.4.5 Action-Level Recording

The Actions view exposes the raw action log recorded for the session. Each entry corresponds to an operation performed within the workflow and is tagged by action type and, where applicable, by the performing AI system. As shown in Figure 4.6, the visible entries include GENERATE and PROVIDE actions displayed in reverse-chronological order.

This view is especially useful because it reveals the operational layer from which the other provenance representations are derived. The session timeline, derivation relationships, and attribution calculations all depend on this recorded sequence of actions. In relation to the EAA model, it provides direct evidence of the Action layer. In relation to the governance requirements, it demonstrates that the system captures workflow granularity at the level of individual operations rather than only coarse summaries of origin.

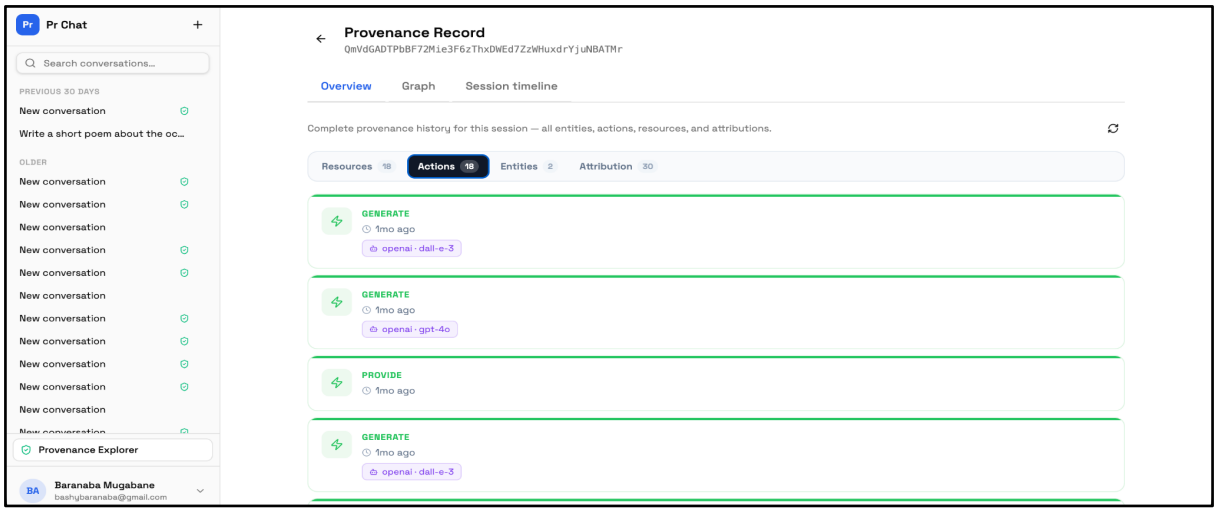


Figure 4.6. Actions view of the provenance record

4.4.6 Typed Participant Representation

The Entities view lists the distinct participants recorded in the provenance bundle. Each entity is displayed with a type tag and an associated identifier. In the illustrated session, shown in Figure 4.7, the recorded entities are the AI model openai/gpt-4o, tagged as AI, and the user's decentralised identifier, tagged as Human.

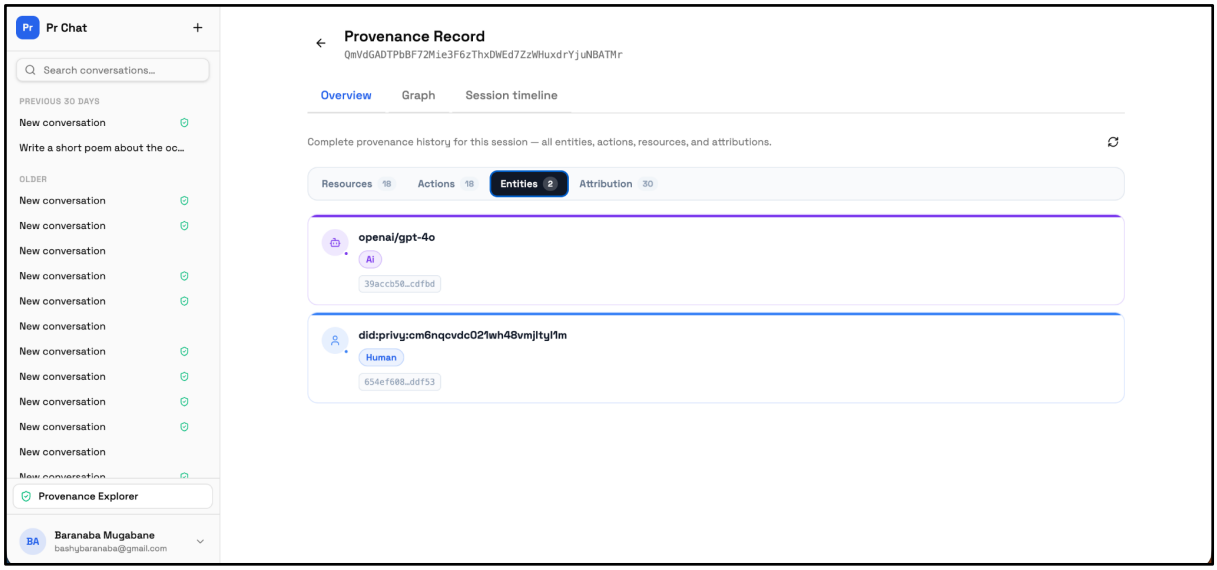


Figure 4.7. Entities view of the provenance record

This view demonstrates the typed distinction between human and AI participants that is central to the EAA model. Rather than inferring participation indirectly from context, the system records participant type as a structural property of the provenance record itself. This directly

supports R1, since the requirement is not only that acts be recorded, but that human and AI acts be recorded as distinct, typed events linked to identified content artefacts.

4.4.7 Attribution Representation

The Attribution view surfaces the computed attribution distribution across the session. In the example shown in Figure 4.8**,** openai/gpt-4o is assigned 80% and the human entity 20%. This distribution is generated from the recorded session using the contribution logic implemented through the attribution extension.

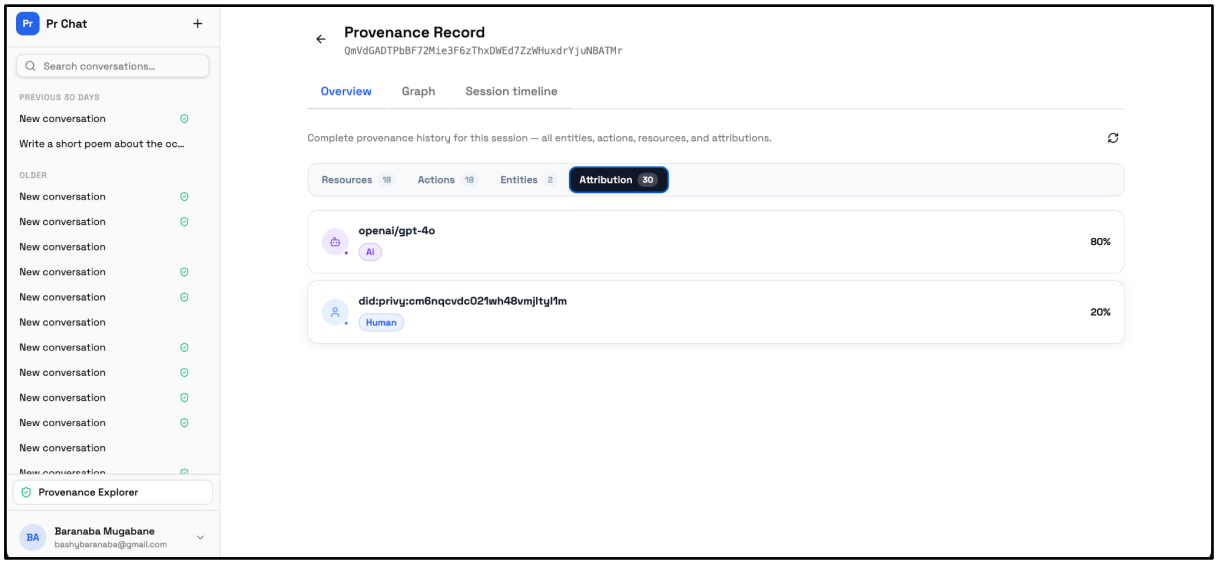


Figure 4.8. Attribution view of the provenance record

This view illustrates an important consequence of modelling attribution as a first-class part of the provenance record rather than as an annotation added only to the final output. Because actions, entities, and resources are recorded across the workflow, the system can compute contribution distributions over the session as a whole. This does not itself determine legal authorship, but it demonstrates how the provenance record can support structured credit allocation and contribution analysis in a reproducible way.

4.5 Evaluation

The evaluation considered whether the approach, as designed and implemented, satisfied the governance requirements derived from the legal and technical analysis, whether it exhibited the expected structural properties of an interoperable and privacy-aware provenance system, and whether the implementation provided supporting evidence of technical operability.

4.5.1 Evaluation Against R1 to R6

All six functional requirements are addressed by the approach as designed. R1 is satisfied by the EAA type system and the per-operation Action record, which distinguish typed human and AI participation at the workflow level. R2 is satisfied by the licensing and authorisation extensions, which encode rights reservation and authorisation conditions as first-order provenance metadata. R3 is satisfied by the use of traversable derivation relations linking transformed outputs to upstream inputs through action-resource chains. R4 is satisfied by the dual-mode AI extension, which records structured AI metadata and distinguishes AI-as-tool from AI-as-agent contexts. R5 is satisfied by the authorisation extension, which records consent-related and authorisation-related metadata at the artefact level. R6 is substantially satisfied by the five-mechanism privacy module. The main qualification is that encrypted vector search requires the verifying party to hold an appropriate decryption key, which constrains fully trustless verification where no such key is available.

4.5.2 Structural Evaluation of the Approach

Beyond direct requirements coverage, the artefact was also evaluated in structural terms. The approach exhibits the core properties expected of a workflow-level provenance system designed for governance use in Human-AI contexts. Data capture is eager and per-operation. Granularity is fine-grained at the workflow level, with explicit distinction between participants, actions, resources, and attributions. Orientation is process-centred, with artefact identity embedded through content-addressed references. Interoperability is supported through W3C PROV alignment, the EAA context URI, and bidirectional C2PA conversion. Storage is pluggable across heterogeneous backends, and privacy is addressed through multiple mechanisms integrated into the architecture rather than added externally. This structural evaluation shows that the approach is not only requirements-complete in a narrow sense, but also coherent as a provenance architecture capable of supporting governance-oriented use across different technical environments.

4.5.3 Supporting Implementation Evidence

The evaluation was further supported by implementation-level evidence in the form of the automated test suite accompanying the operational implementation. At the time of submission, the implementation included 1,099 passing automated tests across smart contracts, storage,

indexing, extensions, payment adapters, the privacy module, git provenance, media and C2PA handling, the SDK, and API integration. These tests provide reproducible evidence of technical operability and specification conformance across the implemented components of the artefact. A summary of the test suite and its package-level coverage is provided in Appendix A2.

4.5.4 Limitations of the Artefact

Five limitations apply. First, no independent regulatory review of legal sufficiency has been conducted; the approach is designed to support compliance, but this has not been assessed by legal practitioners. Second, prototype integrations were developed by the approach's authors and do not constitute independent validation. Third, the `ext:namespace@semver` versioning scheme has no external governance body, representing a risk to long-term semantic interoperability. Fourth, the indexer requires manual operation and configuration. Fifth, at the time no database schema migration tooling is provided for approach version upgrades.

4.6 Summary

The literature review established that no existing system collectively provides workflow-level provenance records that are governance-oriented, rights-reservation-aware, and usable across a decentralised, multi-stakeholder environment while preserving privacy. The approach developed in this study addresses that gap across all identified dimensions. The EAA data model provides typed, workflow-level production records supporting authorship assessment and responsibility attribution. Traversable derivation chains support infringement evaluation. Rights reservation status and authorisation conditions are encoded as first-order metadata rather than as natural-language annotations. Governance-relevant verification without unrestricted data exposure is provided by a layered privacy module. The extension system and pluggable infrastructure support cross-context usability in a multi-stakeholder environment. The chat-based demonstration shows that the approach is deployable in a representative application setting through instrumentation at the application layer.

The approach therefore provides what the literature review showed to be absent: a governance-oriented, workflow-level provenance specification whose informational model is complete enough to support attribution, disclosure, licensing, and responsibility assessment, while also remaining structured, interoperable, and operationally demonstrable within contemporary Human-AI workflows.

Chapter 5: Discussion

The central claim of this study is that governance of Human-AI-created works is, at its foundation, an information problem. Existing legal and regulatory frameworks require answers to questions about how a work was produced, who contributed to it, what materials entered the workflow, and under what conditions relevant acts took place. Yet existing technical systems address these questions only partially and in isolation. The approach developed in this study responds to that deficit by providing a governance-oriented, workflow-level provenance approach designed to make the Human-AI creation processes more legible, traceable, and governable.

5.1 Alignment with Legal, Regulatory, and Technical Literature

RQ1 asked which legal and governance requirements shape attribution, responsibility, liability, and transparency for Human-AI-created works. The Copyright authorship doctrine, the DSM Directive, the EU AI Act, the GDPR, and emerging AI legislations each impose distinct informational obligations that when read together reveal specifications that a provenance approach should meet. The legal analysis that grounds RQ1 showed that authorship and copyrightability in Human-AI works are process-dependent. In both the United States and the European Union, the decisive issue is not merely the existence of an output, but the character and extent of human contribution across the stages through which that output was produced. The present findings are consistent with that position. By recording prompts, generations, selections, edits, and derivation chains, the approach provides the kind of process record that those legal standards presuppose. The DSM Directive requires machine-readable rights reservations if they are to operate effectively at scale, and the EU AI Act requires AI-generated outputs to be marked in machine-readable and interoperable form. The approach shows that these requirements can be represented as first-order provenance metadata within an interoperable model rather than as external annotations or platform-specific flags.

RQ2 required engagement with the technical provenance literature, and the findings refine it in important ways. Pérez et al. (2018) describe provenance as a mature but heterogeneous field and identify interoperability, security/privacy, and verification as among the least supported properties across existing systems. The present study responds directly to those weaknesses. It demonstrates that provenance for Human-AI works must do more than record derivation history. It must also distinguish human and AI participation, carry governance-relevant

metadata, support interoperable representation, and manage the tension between accountability and confidentiality. In this respect, the approach extends the technical provenance literature by showing how legal and regulatory requirements can function as design requirements for provenance modelling.

The study is also consistent with Kale et al.'s (2023) argument that provenance contributes something distinct from post-hoc explainability. Explanation methods may describe model behaviour or provide local reasons for an output, but they do not preserve the wider workflow context through which results emerge. The approach developed here supports that distinction. Its contribution lies not in explaining model internals, but in documenting the process history, metadata, and relationships among artefacts, activities, and actors. This makes provenance complementary to explainability rather than a substitute for it.

5.2 Design Choices and Their Rationale

RQ3 asked how an interoperable and privacy-aware provenance approach can be designed to capture the entities, actions, artefacts, derivations, and governance-relevant conditions involved in Human-AI creation processes. The answer is the layered architecture presented in chapter 4 and three design choices within it are worth examining critically.

The most consequential choice was to derive every component from a specific legal instrument rather than from generic data-management principles. The EAA model's typed entity distinction exists because copyright doctrine requires per-actor differentiation (R1). The aiTraining field in ext:license@1.0.0 exists because DSM Article 4(3) requires machine-readable opt-outs (R2). The dual-mode structure of ext:ai@1.0.0 exists because EU AI Act Article 50(2) distinguishes AI-as-tool from AI-as-agent contexts (R4). This legal grounding means that the approach can claim not just legal relevance but legal traceability: if requirements change because a court reinterprets the authorship standard or the EU AI Act implementing guidance specifies new disclosure formats, the affected component can be identified and revised without redesigning the whole system. This is a structural advantage over approaches designed around generic provenance dimensions.

Comparison with adjacent technical approaches is also instructive. AI detection tools and watermarking mechanisms remain useful for identifying likely AI-generated content or signalling AI involvement, but they remain largely output-oriented. They do not record the wider workflow in which prompts, intermediate outputs, edits, rights conditions, and

authorisation status become legally significant. C2PA goes further by supporting asset-level provenance and edit history, and this study does not treat it as inadequate. Rather, the findings suggest that C2PA and the proposed approach address related layers of the problem. Content authenticity systems are strongest where the asset and its attached credential are the principal concern. The present approach extends beyond that by modelling actor roles, derivation chains, rights reservations, authorisation conditions, and privacy-aware disclosure at the workflow level. In that sense, the study extends rather than rejects the value of existing authenticity and traceability systems.

5.3 Extent of Support and Its Limits

RQ4 asked to what extent the proposed approach can support attribution, disclosure, licensing compliance, and responsibility assessment in representative Human-AI creation scenarios. This is the most important question and the one whose answer must be stated most carefully.

The demonstration results in chapter 4 establish that the approach is deployable in representative platform contexts. The generative AI platform shows that provenance recording can be integrated as a side effect of normal response generation without interrupting user interaction. The findings support the argument that provenance recording should operate at infrastructure level rather than as an explicit user action, a precondition for it to function at platform scale. These findings are directly relevant to RQ4 because they show that the requirements derived from legal analysis can be satisfied in working software in realistic contexts, not only in theory.

The evaluation results establish that all six requirements are addressed as designed. The qualification on R6 is genuine: encrypted vector search requires the verifying party to hold a decryption key, which constrains fully trustless verification in scenarios where a verifier holds no keys. This is an acknowledged boundary on the extent of support.

The extent of support for RQ4 must, however, be assessed against what the evaluation did not include. The choice of Design Science Research Methodology was appropriate for a study whose purpose was to develop and evaluate a working artefact. It allowed normative doctrinal analysis and technical system analysis to be translated into explicit design requirements and then into an operational approach. At the same time, the methodology places limits on what the findings can establish. Scenario-based demonstration with author-built prototypes confirms functional deployment and requirements coverage, but it does not establish scale, adversarial

robustness, or legal sufficiency in real-world proceedings. No independent party deployed the approach against adversarial inputs. No legal practitioner reviewed whether the records produced would satisfy evidentiary standards in actual infringement or regulatory proceedings. No scale testing was conducted. These absences do not invalidate the findings, but they bound them. The correct characterisation of the RQ4 answer is that the approach satisfies the identified requirements at the level of design and functional prototype, and that establishing the extent of support in operational, adversarial, and legal contexts is the primary direction for future work. Similarly, the doctrinal analysis is grounded in authoritative and recent sources, but the legal environment in this area remains dynamic. Future judicial decisions, regulatory guidance, or platform practices may shift the requirements that a provenance approach must satisfy, and the requirements derivation would need to be revisited accordingly.

5.4 Intrinsic Limitations of Provenance as a Governance Mechanism

Further limitations deserve candid emphasis. First, like all provenance systems, the approach depends on cooperative recording. Provenance that is never captured cannot be recovered, regardless of how trustworthy later storage or verification mechanisms may be. Second, the developed approach records what actors and systems claim at the point of capture, not whether those claims are true. Tamper-evidence and non-repudiation can strengthen trust in recorded events after the fact, but they do not prevent false or incomplete metadata from being recorded in the first place. The records are therefore best understood as structured and probative evidence of claims made by identified actors at specific times, not as self-sufficient proof of underlying fact. This limitation is not unique to the present approach, but it sets the correct expectation for legal and governance use.

The privacy dimension must also be understood carefully. The approach manages the tension between accountability and data minimisation; it does not eliminate it. Selective disclosure can allow a creator or platform to prove specific provenance facts without exposing the entire workflow, which is important where prompts, source materials, or intermediate outputs are commercially sensitive or privacy sensitive. This also means that access to the full provenance record remains a governance question in its own right. The approach provides the technical means to implement differentiated access; it does not determine who ought to see what. That remains a matter for law, policy, and institutional design.

5.5 Contributions

Despite these limitations, the study makes several contributions. Conceptually, it reframes provenance for Human-AI works as a governance problem rather than only a data-management problem. Methodologically, it shows how legal and regulatory analysis can be used to derive system requirements and shape provenance modelling. Technically, it proposes a workflow-level approach that integrates artefact history, actor differentiation, AI disclosure, rights metadata, and privacy-aware verification. Practically, it demonstrates that such a provenance approach can be integrated into representative Human-AI environments without requiring a complete redesign of the host system.

Provenance should not be understood as a system that decides authorship, infringement, or liability on its own. Those remain legal judgments. What the approach contributes is the evidentiary and governance infrastructure needed to make those judgments better informed and more traceable. In that respect, the study addresses an infrastructure problem rather than the totality of the trust problem. Overall, the findings suggest that the main challenge in Human-AI governance is not the absence of legal rules or technical tools in isolation, but the absence of an informational infrastructure that connects them. The approach developed here is a response to that gap. Its significance lies in making Human-AI creation workflows more legible and more governable, while remaining clear about the limits of what provenance alone can achieve.

Chapter 6: Conclusion

This thesis examined how a governance-oriented provenance approach for Human-AI-created works can address the informational demands of copyright doctrine, AI disclosure regulation, and data governance. It shows that this can be achieved through workflow-level provenance structures.

6.1 Answers to the Research Questions

The research questions are answered as follows. For RQ1, it showed that attribution, responsibility, liability, and transparency in Human-AI-created works are shaped by combined requirements drawn from copyright doctrine, the DSM Directive, the EU AI Act, the GDPR, and related consent-sensitive frameworks, all of which depend on process-level information. For RQ2, it found that attribution and accountability require five categories of provenance information: participants, artefacts, actions, derivations, and governance-relevant metadata. For RQ3, it proposed an interoperable and privacy-aware operational approach built around a stable provenance core, semantic extensibility, selective disclosure, independent verification, and deployable integration layers. For RQ4, it demonstrated that the proposed approach supports attribution, disclosure, licensing compliance, and responsibility assessment in representative scenarios at the level of design and prototype operation, while remaining subject to limits of validation, scale, and legal sufficiency. These findings fulfil the main aim of the thesis by not only identifying the relevant legal and governance requirements, but also specifying what information they require, how that information can be structured, how it can be implemented operationally, and how far such an approach can support governance in practice.

6.2 Significance and Practical Implications

The implications of this work are practical as well as theoretical. For researchers, the thesis positions provenance for Human-AI creation as a governance problem rather than only a data-management problem. For platform operators and AI developers, the most actionable implication is that EU AI Act Article 50(2) and DSM Directive Article 4(3) compliance is technically tractable today through application-layer instrumentation that does not require fundamental architectural changes. For policymakers and rights holders, the study suggests that machine-readable disclosure and rights-reservation requirements are not merely aspirational

but technically implementable if supported by shared standards and adoption. Conceptually, the thesis suggests that the central gap is not purely legal or purely technical, but one of alignment: current governance frameworks, recording infrastructures, and stakeholder needs do not yet converge on a shared and usable model of process-level information for Human-AI created works.

6.3 Limitations and Constraints

The proposed approach was evaluated through scenario-based demonstrations and prototype implementation rather than independent large-scale deployment, so the findings establish feasibility and functional adequacy, but not adversarial robustness, operational scalability, or legal sufficiency in real-world proceedings. The doctrinal analysis was also limited to selected frameworks, especially those of the European Union, the United States, and the broader international intellectual property context represented by WIPO. In addition, some limits arise from provenance itself as a governance mechanism: unrecorded actions cannot be reconstructed, recorded metadata may still be incomplete or inaccurate, and selective disclosure manages rather than resolves the tension between accountability and restricted access. The approach should therefore be understood as structured governance infrastructure and probative evidence, not as self-sufficient proof of underlying fact.

6.4 Future Works

The limitations of the study point directly to future research. Because the approach was evaluated through author-developed, scenario-based demonstrations, the findings establish design feasibility rather than scale, adversarial robustness, or real-world legal sufficiency. Future work should therefore include independent operational testing and legal evaluation of whether provenance records of this kind satisfy evidentiary standards in contexts such as copyright proceedings, AI Act audits, and GDPR accountability inquiries. Further research is also needed on adaptation across other jurisdictions and future legal developments, as well as on the standardisation and governance of shared provenance vocabularies to avoid fragmentation. Overall, more effective governance of Human-AI-created works will depend on closer convergence between legal requirements, provenance design, and stakeholder-oriented access to trustworthy process information.

References

Altintas, I., Barney, O., & Jaeger-Frank, E. (2006). Provenance collection support in the Kepler scientific workflow system. In L. Moreau & I. Foster (Eds.), Proceedings of the International Provenance and Annotation Workshop (IPAW 2006) (Lecture Notes in Computer Science, Vol. 4145, pp. 118–132). Springer.

Andersen v. Stability AI Ltd, No. 3:23-cv-00201 (N.D. Cal. 2023).

Authors Guild v. Google, Inc., 804 F.3d 202 (2d Cir. 2015).

Berne Convention for the Protection of Literary and Artistic Works, Sept. 9, 1886, as revised at Paris on July 24, 1971, S. Treaty Doc. No. 99-27 (1979).

C2PA. (2026). C2PA technical specification, version 2.3. Coalition for Content Provenance and Authenticity. https://spec.c2pa.org/specifications/specifications/2.3/specs/C2PA_Specification.html

Campbell v. Acuff-Rose Music, Inc., 510 U.S. 569 (1994).

Community for Creative Non-Violence v. Reid, 490 U.S. 730 (1989).

Content Authenticity Initiative. (n.d.). How it works. https://contentauthenticity.org/how-it-works

Council of Europe. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union, L 2024/1689.

Downing v. Abercrombie & Fitch, 265 F.3d 994 (9th Cir. 2001).

European Parliament and Council of the European Union. (1996). Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases. Official Journal of the European Union, L 77, 20–28.

European Parliament and Council of the European Union. (2001). Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society. Official Journal of the European Union, L 167, 10–19.

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Official Journal of the European Union, L 119, 1–88.

European Parliament and Council of the European Union. (2019). Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. Official Journal of the European Union, L 130, 92–119.

European Parliament and Council of the European Union. (2022). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a single market for digital services and amending Directive 2000/31/EC (Digital Services Act). Official Journal of the European Union, L 277, 1–102.

Federal Communications Commission. (2024). Declaratory ruling on implications of artificial intelligence technologies on protecting consumers from unwanted robocalls and robotexts, CG Docket No. 23-362.

Feist Publications, Inc. v. Rural Telephone Service Co., 499 U.S. 340 (1991).

Football Dataco Ltd v Yahoo! UK Ltd, C-604/10, EU:C:2012:115 (Court of Justice of the European Union, 2012).

Ganescu, B. M., & Passerat-Palmbach, J. (2024). Trust the process: Zero-knowledge machine learning to enhance trust in generative AI interactions. arXiv preprint arXiv:2402.06414. https://doi.org/10.48550/arXiv.2402.06414

Getty Images (US), Inc. v. Stability AI, Ltd., No. 1:23-cv-00135 (D. Del. 2023).

Ginsburg, J. C., & Budiardjo, L. A. (2019). Authors and machines. Berkeley Technology Law Journal, 34(2), 343–448.

Harper & Row Publishers, Inc. v. Nation Enterprises, 471 U.S. 539 (1985).

Infopaq International A/S v Danske Dagblades Forening, C-5/08, EU:C:2009:465 (Court of Justice of the European Union, 2009).

Institute of Electrical and Electronics Engineers. (2023). IEEE Std 2952-2023: IEEE standard for secure computing based on trusted execution environment. IEEE.

Midler v. Ford Motor Co., 849 F.2d 460 (9th Cir. 1988).

Moreau, L., Clifford, B., Freire, J., Futrelle, J., Gil, Y., Groth, P., Kwasnikowska, N., Miles, S., Missier, P., Myers, J., Plale, B., Simmhan, Y., Stephan, E., & Van den Bussche, J. (2011). The Open Provenance Model core specification (v1.1). Future Generation Computer Systems, 27(6), 743–756. https://doi.org/10.1016/j.future.2010.07.005

Muniswamy-Reddy, K. K., Holland, D. A., Braun, U., & Seltzer, M. I. (2006). Provenance-aware storage systems. In Proceedings of the USENIX Annual Technical Conference 2006 (USENIX ATC '06) (pp. 43–56). USENIX Association.

Nakamoto, S. (2008). Bitcoin: A peer-to-peer electronic cash system. https://bitcoin.org/bitcoin.pdf

Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2024, H.R. 9551, 118th Cong. (2024). https://www.congress.gov/bill/118th-congress/house-bill/9551

Nurture Originals, Foster Art, and Keep Entertainment Safe Act of 2025, S. 1367, 119th Cong. (2025). https://www.congress.gov/bill/119th-congress/senate-bill/1367

Omogiate, O. (2024). Artificial intelligence and intellectual property: Rethinking authorship in the age of generative AI. Journal of Intellectual Property Law and Practice, 19(3), 210–228.

Painer v Standard VerlagsGmbH, C-145/10, EU:C:2011:798 (Court of Justice of the European Union, 2011).

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007–2008). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Pereira, G., Chaari, M. Z., & Daroge, F. (2025). IoT-enabled smart drip irrigation system using ESP32. IoT, 4, 221–243. (Referenced for EU AI Act Article 12 record-keeping context.)

Pérez, B., Rubio, J., & Sáenz-Adán, C. (2018). A systematic review of provenance systems. Knowledge and Information Systems, 57(3), 495–543. https://doi.org/10.1007/s10115-018-1164-3

Ranganthan, V. P., Dantu, R., & Paul, A. (2024). Blockchain-powered data provenance for AI model audits. Scientific Journal of Artificial Intelligence and Blockchain Technologies, 1(1), 28–36. https://doi.org/10.63345/sjaibt.v1.i1.104

Samuelson, P., & Springman, C. J. (2023). Is training generative AI on copyrighted works fair use? Communications of the ACM, 66(9), 21–24.

Samuelson, P., Sag, M., & Springman, C. J. (2023). Brief of intellectual property and technology law professors as amici curiae in support of neither party. In Andersen v. Stability AI Ltd, No. 3:23-cv-00201 (N.D. Cal. 2023).

Thaler v. Perlmutter, 687 F. Supp. 3d 140 (D.D.C. 2023).

Thaler v. Perlmutter, No. 23-5233 (D.C. Cir. 2025).

U.S. Copyright Act of 1976, 17 U.S.C. §§ 101–1401.

U.S. Copyright Office. (2024). Copyright and artificial intelligence, Part 1: Digital replicas. United States Copyright Office.

U.S. Copyright Office. (2025). Copyright and artificial intelligence, Part 2: Copyrightability. United States Copyright Office.

Weber-Wulff, D., Anohina-Naumeca, A., Bjelobaba, S., Foltýnek, T., Guerrero-Dib, J., Popoola, O., Šigut, P., & Waddington, L. (2023). Testing of detection tools for AI-generated text. *International Journal for Educational Integrity, 19*, 26. https://doi.org/10.1007/s40979-023-00146-z

Wolstencroft, K., Haines, R., Fellows, D., Williams, A., Withers, D., Owen, S., Soiland-Reyes, S., Dunlop, I., Nenadic, A., Fisher, P., Bhagat, J., Belhajjame, K., Bacall, F., Hardisty, A., Nieva de la Hidalga, A., Balcazar Vargas, M. P., Sufi, S., & Goble, C. (2013). The Taverna workflow suite: Designing and executing workflows of Web Services on the desktop, web or in the cloud. Nucleic Acids Research, 41(W1), W557–W561. https://doi.org/10.1093/nar/gkt328

World Intellectual Property Organization. (2020). WIPO Intellectual Property Handbook. WIPO Publication No. 489(E). World Intellectual Property Organization.

World Intellectual Property Organization. (n.d.). Understanding copyright and related rights. World Intellectual Property Organization. Retrieved February 2026, from https://www.wipo.int/copyright/en/

World Wide Web Consortium. (2013a). PROV-overview: An overview of the PROV family of documents (W3C Working Group Note, 30 April 2013). https://www.w3.org/TR/prov-overview/

World Wide Web Consortium. (2013b). PROV-DM: The PROV data model (W3C Recommendation, 30 April 2013). https://www.w3.org/TR/prov-dm/

World Wide Web Consortium. (2013c). PROV-O: The PROV ontology (W3C Recommendation, 30 April 2013). https://www.w3.org/TR/prov-o/

Xing, Z., Liu, Y., Li, X., Wang, Y., Zhang, L., Li, B., & Mao, B. (2023). Zero-knowledge proof meets machine learning in verifiability: A survey. arXiv preprint arXiv:2310.14848. https://doi.org/10.48550/arXiv.2310.14848

Appendices

Appendix A1: Gantt Chart



Appendix A2: Test Suite Summary

Package	Tests	Scope
`@provenancekit/contracts`	27	Smart contract behaviour (Foundry/Solidity): entity registration, resource registration, action recording, proof and commitment layer, hook dispatch
`@provenancekit/storage`	93	All four database adapters (in-memory, PostgreSQL, MongoDB, Supabase) and all six file-storage adapters (IPFS-Pinata, IPFS-Infura, IPFS-local, Web3.Storage, Arweave, memory)
`@provenancekit/indexer`	66	Event transformation functions (5 event types) and ProvenanceIndexer class: historical sync and real-time polling
`@provenancekit/extensions`	377	All 15 ext:namespace@semver extensions including ext:ai@1.0.0 (tool and agent modes), ext:license@1.0.0, ext:authorization@1.0.0, ext:contrib@1.0.0, and ext:x402@1.0.0
`@provenancekit/payments`	43	Three payment adapters: Direct, 0xSplits, and Superfluid
`@provenancekit/privacy`	187	Five privacy modules: encryption-at-rest (XChaCha20-Poly1305), selective disclosure, Pedersen commitments, token-gated access control, encrypted vector search
`@provenancekit/git`	71	Git provenance recording and commit attribution
`@provenancekit/media`	157	C2PA integration, media provenance recording, and bidirectional bundle conversion
`@provenancekit/sdk`	47	IChainAdapter interface, createViemAdapter, createEIP1193Adapter, and the pk.file() integration path
`provenancekit-api`	31	API integration tests covering provenance recording, session management, bundle retrieval, rate limiting (sliding window, 429), and management routes
Total	**1,099**	**All tests passing**