

RESEARCH ARTICLE

Transparency Without Exposure: A Selective Verifiability Framework for Copyright and Trade Secrets in Generative AI Training Data

Heungsok Cha

Professor, Department of Global Business and Trade, Joongbu University, Republic of Korea

Corresponding Author: Heungsok Cha (heungsok@joongbu.ac.kr)

ABSTRACT

Governance of generative artificial intelligence (AI) faces a structural dilemma. Copyright owners and regulators need evidence of training-data acquisition, licensing, rights reservations, and processing, while unrestricted disclosure may expose valuable dataset compilations and methods. This article systematically compares AI transparency duties in Korea, the European Union, the United States/California, and Japan by examining disclosure and information-provision requirements, labeling, obligated actors and triggers, exceptions and safeguards, and enforcement and sanctions. Copyright fair use and text-and-data-mining limitations are treated separately: they govern specified uses of protected subject matter and do not generally displace independent transparency duties covering copyrighted and non-copyright information. The comparison finds that public summaries improve orientation but generally cannot establish whether a particular work was included, on what legal basis it was processed, or in which model version it was used. Therefore, the article proposes “selective verifiability”: the capacity for an authorized actor to test a bounded compliance claim against lifecycle-linked records without receiving unrestricted corpus access. The framework has three tiers: a standardized public training-content summary; a confidential, source-level provenance manifest available to regulators or accredited auditors; and a claim-triggered verification channel for rights holders. W3C PROV, SPDX, machine-learning bills of materials, digital signatures, and append-only commitments can support the architecture but cannot independently prove legal validity or record completeness. For Korea, the article recommends a phased pilot, explicit institutional allocation, purpose-limited data processing, proportionate recordkeeping duties, reviewable trade-secret claims, and evidentiary incentives that do not immunize infringement.

Open Access

Received: July 11, 2026**Revised:** August 06, 2026**Accepted:** September 04, 2026**Published:** September 30, 2026

Funding: The author received manuscript fees for this article from Korea Institute of Intellectual Property.

Conflict of interest: No potential conflict of interest relevant to this article was reported.

© 2026 Korea Institute of Intellectual Property



This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 (<https://creativecommons.org/licenses/by-nc-nd/4.0/>) which permits use, distribution and reproduction in any medium, provided that the article is properly cited, the use is non-commercial and no modifications or adaptations are made.

KEYWORDS

generative AI, training data, copyright, trade secrets, transparency, data provenance, selective verifiability

원저

노출 없는 투명성: 생성형 AI 학습데이터의 저작권 및 영업비밀을 위한 선택적 검증가능성 프레임워크

차형석

중부대학교 글로벌무역(국제통상)학전공 교수

교신저자: 차형석 (heungsok@joongbu.ac.kr)

차례

1. Introduction
 - 1.1. Scope and comparative method
2. From Transparency to Selective Verifiability
 - 2.1. The training-data evidence problem
 - 2.2. Trade secrecy as a legitimate but bounded constraint
 - 2.3. Transparency, traceability, auditability, and verifiability
3. Comparative Analysis of AI Transparency Obligations
 - 3.1. Korea
 - 3.2. European Union
 - 3.3. United States and California
 - 3.4. Japan
 - 3.5. Comparative findings
4. A Three-Tier Selective Verifiability Framework
 - 4.1. Design principles
 - 4.2. Tier 1: public training-content summary
 - 4.3. Tier 2: confidential provenance manifest
 - 4.4. Tier 3: claim-triggered verification
 - 4.5. Legal safeguards, proof allocation, and data protection
 - 4.6. Technical implementation and evidentiary limits
5. A Policy Path for Korea
 - 5.1. Relationship to Article 35-5 and substantive legality
 - 5.2. Legal form and institutional allocation
 - 5.3. Proportional obligations and incentives
 - 5.4. Phased implementation
6. Conclusion

국문초록

생성형 AI 학습데이터의 투명성은 저작권자가 권리를 행사하기 위한 중요한 전제이지만, 데이터 구성·수집·정제 방식의 전면 공개는 개발자의 영업비밀과 경쟁력을 훼손할 수 있다. 본 연구는 한국, 유럽연합, 미국(캘리포니아), 일본의 AI 투명성 규범을 정보 공개·제공 및 표시 의무, 적용 대상과 예외, 책임과 제재라는 공통 기준으로 비교한다. 이들 의무는 저작물 이용의 적법성을 판단하는 공정이용 또는 텍스트·데이터 마이닝(TDM) 제한규정과 규율 대상이 다르며, 학습정보에는 저작권으로 보호되지 않는 데이터도 포함된다. 비교 결과를 토대로 공개 자체보다 적법성 검증에 초점을 둔 '선택적 검증가능성' 체계를 제안한다. 이 체계는 ① 출처·유형·권리상태를 범주화한 공개 요약, ② 규제기관 또는 독립감사인이 비공개로 열람하는 항목·출처 수준의 출처명세, ③ 권리자의 구체적 청구에 한정된 포함 여부 확인 절차로 구성된다. W3C PROV, SPDX 및 ML-BOM과 같은 기계가독 표준과 무결성 통제를 결합하면 원 자료를 공개하지 않고도 수집경로, 이용허락, 권리유보 준수 및 처리이력의 검증을 지원할 수 있다. 한국에는 개인정보 최소화, 검토 가능한 영업비밀 주장, 명확한 증거효과, 중소기업 비례성 및 단계적 이행을 결합한 제도 도입을 제안한다.

주제어

생성형 인공지능, 학습데이터, 저작권, 영업비밀, 투명성, 데이터 출처이력, 선택적 검증가능성

1. Introduction

Generative-AI governance is increasingly an evidence problem. Rights holders cannot assess whether works were collected, licensed, excluded, or retained, and regulators cannot test acquisition controls or rights reservations from model outputs alone. Yet the most probative material, source inventories, licenses, collection windows, filtering rules, and dataset mixtures, may reveal commercially valuable choices. Even as disclosure practices become more formalized, major developers remain opaque about upstream data and development processes.¹⁾

The relevant choice is therefore not between complete disclosure and secrecy. It is between indiscriminate exposure and a safeguarded process in which an authorized actor can test a defined claim. This matters because the legality of training remains fact-specific: Korea's 2026 fair-use guide, for example, identifies source, access, license scope, rights signals, purpose, output controls, and market effects rather than supplying a categorical answer.²⁾

Existing documentation is often insufficient: a large dataset audit found extensive license omissions and errors, while public summaries remain useful only to the extent that they are specific and legally meaningful. Trade secrecy nevertheless protects genuine value in dataset selection, cleaning, weighting, and combination. The institutional challenge is to make evidence available for legitimate scrutiny without publicly revealing an entire corpus or pipeline.³⁾⁴⁾⁵⁾

This article asks how law can enable credible verification of training-data use without exposing commercially sensitive datasets and methods. It contributes a relational account of transparency, traceability, auditability, and verifiability; compares the functions supplied by Korean, EU, United States/California, and Japanese law; and integrates three access routes, a public summary, a confidential lifecycle-linked manifest, and bounded claimant verification, with trade-secret review, data minimization, and procedural safeguards. The contribution is the institutional integration of these mechanisms, not a claim that each element is individually novel.

Selective verifiability neither determines whether training is lawful nor creates a copyright exception. It improves the evidence through which existing standards are

- 1) Alexander Wan et al., "The 2025 Foundation Model Transparency Index", arXiv, <<https://arxiv.org/abs/2512.10169>>, 작성일: 2025. 12. 11.; Stanford Center for Research on Foundation Models, "Foundation Model Transparency Index", Stanford Center for Research on Foundation Models, <<https://crfm.stanford.edu/fmti/December-2025/index.html>>, 검색일: 2026. 7. 1.
- 2) Ministry of Culture, Sports and Tourism & Korea Copyright Commission, "Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models", Korea Copyright Commission, 2026, pp. 1-75.
- 3) Shayne Longpre et al., "A Large-Scale Audit of Dataset Licensing and Attribution in AI", *Nature Machine Intelligence*, Vol.6(2024), pp. 975-987.
- 4) Adam Buick, "Copyright and AI Training Data—Transparency to the Rescue?", *Journal of Intellectual Property Law & Practice*, Vol.20 No.3(2025), pp. 182-192.
- 5) Perry Keller & Tanya Aplin, "Reconciling Trade Secrets and AI Public Transparency", *The Cambridge Law Journal*, Vol.85 No.1(2026), pp. 151-185.

applied. Cryptographic integrity likewise proves only that a statement existed at a time, not ownership, license validity, fair use, completeness, or truth; legal and institutional review remain necessary.

The analysis covers general-purpose and generative-model training and material updates. Personal-data protection is addressed only where it affects provenance and claimant verification; output copyrightability, patents, personality rights, competition, and comprehensive sectoral privacy compliance are outside scope. Materials are current through 21 August 2026.

1.1. Scope and comparative method

The article uses a structured functional comparison. Korea is the target jurisdiction; the EU differentiates deployer information, public labeling, and GPAI documentation; the United States lacks a single cross-sector federal statute, but California imposes distinct dataset, provenance, and frontier-model duties; and Japan links a promotion statute to guidance rather than EU-style mandatory disclosure. These jurisdictions therefore represent materially different transparency architectures.

Primary sources are classified by legal force and compared through common questions: what must be disclosed, provided, or displayed; which actor, system, model, or use triggers the duty; who receives the information and at what detail; which exclusions, confidentiality safeguards, or equivalent-compliance rules apply; and what enforcement or sanction follows. Copyright limitations are considered separately because they govern specified uses of protected subject matter, not the existence of an independent transparency duty.

The comparison assesses formal legal capacity, not the quality of compliance in practice. The proposed design therefore requires empirical testing with providers, rights holders, auditors, and Korean institutions.

2. From Transparency to Selective Verifiability

2.1. The training-data evidence problem

AI training may involve collection, intermediate copying, extraction, normalization, filtering, tokenization, training, evaluation, and retention. Inputs include copyrighted works and protected databases, but also facts, public-domain material, synthetic data, metadata, and other unprotected information. Copyright limitations, Korea's Article 35-5, United States fair use, EU TDM exceptions, and Japan's non-enjoyment exception, govern specified acts involving protected subject matter. AI transparency rules instead govern disclosure, notice, labeling, documentation, or reporting; compliance with one body of law neither excuses nor establishes compliance with the other.⁶⁾

The evidence problem has three dimensions: identity, or what source or version

entered the pipeline; authority, or which license, exception, or other basis was recorded; and continuity, or whether the source can be linked through transformations and dataset versions to a model. Broad source categories address these questions only weakly.

Better records do not presume infringement. They can document lawful purchase, license scope, respect for a rights reservation, exclusion before training, or separation between research and commercial corpora. Provenance therefore reduces both false allegations and information asymmetry.⁷⁾

2.2. Trade secrecy as a legitimate but bounded constraint

Trade-secret law protects information that is secret, valuable because of secrecy, and reasonably protected. A compilation may qualify even when individual components are public, so a training-data regime must distinguish public facts from confidential value created by selection, weighting, cleaning, and combination.⁸⁾

Secrecy is not an automatic veto. A provider must identify the information, its competitive value, and protective measures. *X.AI LLC v. Bonta* illustrates the point: the court accepted that datasets or processing details could qualify but found generalized assertions insufficient on the preliminary record. The decision was interlocutory, but its narrower lesson is that secrecy claims require particularization.⁹⁾

Confidential disclosure to a regulator, court, or auditor need not destroy secrecy. Protected access can combine statutory confidentiality, necessity, proportionality, logging, and review of contested designations. The design question is who receives what information, for which purpose, and under which safeguards.¹⁰⁾

6) Copyright Act (Republic of Korea), Art. 35-5; Copyright Act of the United States, 17 U.S.C. § 107; 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, OJ L 130, 17.5.2019, Arts. 3-4; Copyright Act (Japan), Act No. 48 of 1970, Art. 30-4; see also Seongwon Kim et al., “An Exploratory Study on New Provisions for Exemption from the Use of Copyrighted Works in Text and Data Mining Processes—Applying to Generative Adversarial Networks”, *Journal of Intellectual Property*, Vol.18 No.1(2023), pp. 147-180; Jeonghun Park, “A Study on the Interpretation and Legislation of Data-Related Laws”, *Journal of Intellectual Property*, Vol.17 No.2(2022), pp. 213-258.

7) United States Copyright Office, “Copyright and Artificial Intelligence, Part 3: Generative AI Training”, United States Copyright Office, 2025, pp. 1-108; Peter Henderson et al., “Foundation Models and Fair Use”, *Journal of Machine Learning Research*, Vol.24 No.400(2023), pp. 1-79; Mark A. Lemley & Bryan Casey, “Fair Learning”, *Texas Law Review*, Vol.99 No.4(2021), pp. 743-785; Pamela Samuelson, “Generative AI Meets Copyright”, *Science*, Vol.381 No.6654(2023), pp. 158-161.

8) World Intellectual Property Organization, “WIPO Guide to Trade Secrets and Innovation”, WIPO, 2024, pp. 1-149; see also Insu Yang & Sunha Park, “A Study on Requirement of Non-Publicity for Trade Secrets”, *Journal of Intellectual Property*, Vol.20 No.1(2025), pp. 1-38.

9) *X.AI LLC v. Bonta*, No. CV 25-12295-JGB (SSCx), Order Denying Preliminary Injunction, pp. 5-6 (C.D. Cal. 4 March 2026).

10) Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, OJ L 157, 15.6.2016, Arts. 3(2), 9.

2.3. Transparency, traceability, auditability, and verifiability

Transparency, traceability, auditability, and verifiability are complementary but distinct. Transparency makes intelligible information available to a defined audience. Traceability reconstructs relationships from a source through transformations and dataset versions to a training run. Auditability permits an authorized independent reviewer to obtain evidence and apply criteria. Verifiability permits a bounded factual or compliance claim to be supported, contradicted, or qualified. A public statement may therefore be transparent but untraceable, and an inspection power may yield little verification when records are incomplete.

Datasheets can improve transparency and W3C PROV can represent lineage, but neither alone lets a copyright owner test whether a specified edition was linked to a model version and on what recorded basis.¹¹⁾¹²⁾

Selective verifiability is the capacity for an authorized actor to test a sufficiently specified claim against contemporaneous, lifecycle-linked records without unrestricted corpus access. Access varies by role and purpose: the public receives categories, a regulator or accredited auditor receives protected evidence, and a verified claimant receives only a bounded work-specific result. It makes transparency actionable rather than replacing it.

Existing law supplies only parts of this design. Korea's Articles 31 and 34 require user notice, output identification, and limited high-impact explanation, not a source-level ledger or claimant right. EU Article 53 provides a public training summary and regulator- or downstream-facing documentation, but no ordinary claimant procedure. Copyright fair use or TDM limitations cannot serve as a general defense to these independent duties, which may concern protected and unprotected information alike.

3. Comparative Analysis of AI Transparency Obligations

Transparency duties differ by function and recipient: operational instructions, interaction notices, synthetic-content labels, training summaries, explanations, and regulator reports are not interchangeable. Each jurisdiction is therefore compared by duty, actor and trigger, recipient, exception or safeguard, and enforcement. Copyright limitations are not treated as transparency exemptions.

3.1. Korea

Korea's AI Basic Act applies to AI developers and businesses using another party's AI to provide products or services, including foreign conduct affecting the Korean

11) Timnit Gebru et al., "Datasheets for Datasets", *Communications of the ACM*, Vol.64 No.12(2021), pp. 86-92.

12) World Wide Web Consortium, "PROV-DM: The PROV Data Model", W3C, <<https://www.w3.org/TR/prov-dm/>>, 작성일: 2013. 4. 30.

market, but excludes designated defense- or security-only systems. Article 31 requires advance notice for services using high-impact or generative AI and labeling of generative outputs and realistic synthetic media. Decree Article 23 permits human- or machine-readable methods and relaxes duties where AI use is obvious, internal only, or within ministerially designated categories.¹³⁾¹⁴⁾

Article 34 separately requires providers of high-impact AI to establish risk, user-protection, oversight, documentation, and technically feasible explanation measures, including an overview of training data. Decree Article 27 requires public posting of the principal measures and five-year retention of supporting documents but permits statutory trade secrets to be withheld. Upstream implementation without material downstream modification and equivalent measures under other laws may satisfy specified duties. This is an explanation framework, not disclosure of each training item or rights-clearance record.¹⁵⁾

The Ministry of Science and ICT may investigate specified labeling and high-impact obligations and order suspension or correction. Failure to give advance notice is directly subject to an administrative fine capped at KRW 30 million; other breaches generally become fineable when an order is disobeyed. No special criminal penalty or private damages action is created. Copyright Article 35-5 remains relevant to protected works but is neither a transparency exemption nor a rule for unprotected data.¹⁶⁾

3.2. European Union

The EU AI Act separates high-risk operational transparency, public notice and labeling, and GPAI-model documentation. Article 13 requires high-risk providers to give deployers instructions on purpose, performance, limits, risks, oversight, logs, and relevant data information, while Article 26 adds specified worker and affected-person notices. Under Regulation (EU) 2026/1744, these high-risk duties apply from December 2027 for Annex III systems and August 2028 for Annex I product systems.¹⁷⁾

Article 50, applicable from 2 August 2026, requires disclosure of direct AI interaction unless obvious, machine-readable marking of synthetic content, notice for emotion recognition or biometric categorization, and disclosure of deepfakes

13) Framework Act on the Development of Artificial Intelligence and the Creation of a Foundation for Trust (Republic of Korea), Act No. 21311 (consolidated version effective 21 July 2026), Arts. 2(7), 4, 31, 34 (hereinafter “AI Basic Act”); Enforcement Decree of the same Act, Presidential Decree No. 36580 (effective 20 August 2026), Arts. 2, 23, 27 (hereinafter “Enforcement Decree”).

14) AI Basic Act, Art. 31; Enforcement Decree, Art. 23; Ministry of Science and ICT (Republic of Korea), “Guidelines for Securing AI Transparency”, Ministry of Science and ICT, 2026, pp. 1–28.

15) AI Basic Act, Art. 34; Enforcement Decree, Art. 27; Act on Prevention of Unfair Competition and Protection of Trade Secrets (Republic of Korea), Art. 2(2).

16) AI Basic Act, Arts. 40, 43; Enforcement Decree, attached Table 2.

17) Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence, OJ L, 12.7.2024, as amended by Regulation (EU) 2026/1744 of 8 July 2026 (Digital Omnibus on AI), OJ L, 24.7.2026, Arts. 13, 26.

and certain public-interest text. Exceptions or adjustments cover authorized law enforcement, standard editing, human-reviewed text with editorial responsibility, and artistic or analogous works. Legacy systems receive the specified transitional period.¹⁸⁾

Article 53 requires GPAI providers to maintain regulator-facing technical documentation, provide downstream information, adopt a Union-copyright compliance policy, and publish an AI Office-template summary of training content. The open-source exception covers only the first two duties and not systemic-risk models. The public summary is broader than copyrighted works: Article 78 protects information received by authorities, while the template balances public detail against confidential business information.¹⁹⁾

Article 50 breaches can attract fines up to EUR 15 million or 3 per cent of worldwide turnover, subject to the Regulation's undertaking and SME rules; the Commission can impose the same ceiling on GPAI providers under Article 101 and order compliance, restriction, withdrawal, or recall. DSM Directive TDM exceptions govern reproduction or extraction of protected material, but do not waive Articles 13, 50, or 53. Article 53 copyright policy and public training summary are independent obligations.²⁰⁾²¹⁾

3.3. United States and California

The United States has no single cross-sector federal AI transparency statute. The FTC Act can address unfair or deceptive representations and sectoral law can regulate particular uses, but federal copyright fair use separately governs uses of protected works. The principal horizontal duties examined here arise under distinct California statutes.²²⁾

California AB 2013 requires a developer, before a covered generative-AI release in California, to post a high-level training-dataset summary addressing source or owner, purpose, volume and type, protected or public-domain content, purchase or licensing, consumer information, processing, timing, and synthetic data. Security, aircraft, and federal defense uses are excluded. The statute has no dedicated enforcement authority, penalty, or private action, and its scope extends beyond copyrighted material.²³⁾

18) Regulation (EU) 2024/1689, Art. 50; European Commission, "Transparency Obligations under Article 50 of the AI Act", European Commission, <<https://digital-strategy.ec.europa.eu/en/faqs/transparency-obligations-under-article-50-ai-act>>, 검색일: 2026. 7. 1.

19) Regulation (EU) 2024/1689, Art. 53, Recitals 106-108, Annexes XI-XII; European Commission, "Explanatory Notice and Template for the Public Summary of Training Content for general-purpose AI models", European Commission, <<https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models>>, 검색일: 2026. 7. 1.

20) Regulation (EU) 2024/1689, Arts. 78, 93, 99, 101.

21) Directive (EU) 2019/790, Arts. 3-4; Regulation (EU) 2024/1689, Art. 53(1)(c)-(d).

22) Federal Trade Commission Act, 15 U.S.C. § 45; Copyright Act of the United States, 17 U.S.C. § 107.

23) California Assembly Bill 2013, Chapter 817 (2024) (Artificial Intelligence Training Data

The California AI Transparency Act applies from 2 August 2026 to publicly accessible generative-AI providers exceeding one million monthly users or visitors. For image, video, and audio, providers must offer a detection tool, a user-selectable manifest disclosure, and technically feasible latent provenance information. Security restrictions, privacy limits, licensed-system rules, and specified entertainment exclusions apply. Public authorities may seek USD 5,000 per violation per day, and some license breaches support injunctions.²⁴⁾

SB 53 requires threshold frontier-model transparency reports and, for large developers, a public risk framework and specified assessment and incident reporting. Reasoned redaction protects trade secrets, security, and legal compliance, with an unredacted version retained for five years; equivalent federal compliance may be recognized. The Attorney General may seek up to USD 1 million per specified violation. Fair use is not an exemption from any of these dataset, provenance, or risk-reporting duties.²⁵⁾

3.4. Japan

Japan's 2025 AI Promotion Act makes process transparency a basic principle and asks AI research, development, and utilization businesses to endeavor to use AI actively and cooperate with government measures. It does not directly require training-data disclosure, user notice, or generated-content labeling.²⁶⁾

Instead, Articles 13 and 16 support national guidelines, fact-finding, guidance, and advice. The December 2025 guidelines recommend information on system limits, data-collection policy, output reliability, source indications, responses to disclosure requests, and watermarking or provenance, while accommodating trade secrets, intellectual property, feasibility, and burden. The Act creates no standalone transparency fine or private action. Copyright Article 30-4 separately addresses non-enjoyment uses and is not a transparency defense.²⁷⁾²⁸⁾

Transparency), Cal. Civ. Code §§ 3110-3111 (operative 1 January 2026).

24) California Senate Bill 942, Chapter 291 (2024) (California AI Transparency Act), as amended by Assembly Bill 853, Chapter 674 (2025), Cal. Bus. & Prof. Code §§ 22757-22757.6 (operative 2 August 2026).

25) California Senate Bill 53, Chapter 239 (2025) (Transparency in Frontier Artificial Intelligence Act), Cal. Bus. & Prof. Code §§ 22757.10-22757.16.

26) Act on the Promotion of Research and Development and Utilization of Artificial Intelligence-related Technology (Japan), Act No. 53 of 2025, Arts. 3(4), 7 (hereinafter "AI Promotion Act").

27) Id., Arts. 13, 16; AI Strategy Headquarters (Japan), "Guidelines on Ensuring Appropriateness in the Research, Development and Utilization of Artificial Intelligence-related Technology", AI Strategy Headquarters, 2025, pp. 1-12.

28) Copyright Act (Japan), supra note 6, Art. 30-4; Agency for Cultural Affairs (Japan), "General Understanding on AI and Copyright in Japan—Overview", Agency for Cultural Affairs, 2024, pp. 1-19.

3.5. Comparative findings

Table 1 summarizes the compared duties, scope, safeguards, and consequences.

<Table 1 Comparative structure of AI transparency obligations>

Function	Korea	European Union	United States / California	Japan
Law/actor	Binding Act/Decree; AI providers	Binding AI Act; high-risk/GPAI actors	Federal mosaic; California developers/providers	Promotion Act plus guidance
Notice	Advance notice for high-impact/generative AI	AI interaction and affected-person notices	No general horizontal duty examined	User information recommended
Content marking	Generated and realistic synthetic content	Machine-readable marking; deepfake/public-interest disclosure	CA provenance for image/video/audio	Watermark/provenance recommended
Training/model info	High-impact explanation may include data overview	GPAI public summary and protected documentation	AB 2013 dataset summary; SB 53 model/risk reports	Collection/source information recommended
Exceptions/safeguards	Obvious/internal use; security; trade secrets; equivalence	Editing, law enforcement, art, limited open source; confidentiality	Security, thresholds, feasibility, redaction, equivalence	Guidance weighs secrets, IP, feasibility, burden
Enforcement	Investigation, orders, fines up to KRW 30m	Up to EUR 15m/3%; Commission GPAI powers	AB 2013 none specified; USD 5k/day or up to USD 1m	Guidance/advice; no standalone sanction
Copyright relation	Art. 35-5 is not an AI Act exemption	DSM TDM does not waive Arts. 13, 50, 53	Fair use is not a California transparency exemption	Art. 30-4 is separate from transparency guidance

Source: Author's classification as of 21 August 2026. Future application dates are identified; guidance and copyright limitations are not coded as binding transparency duties or exemptions.

Four conclusions follow. First, transparency is function- and recipient-specific. Second, only EU Article 53 and California AB 2013 require broadly framed public training-content summaries; Korea requires user notice and limited high-impact explanation, while Japan relies on guidance. Third, exceptions arise from the transparency rules themselves, such as obviousness, security, thresholds, feasibility, open-source limits, equivalence, or protected redaction, not from copyright permission. Fourth, sanctions range from substantial EU and California penalties to Korean corrective orders and fines, AB 2013's lack of a dedicated penalty, and Japan's nonbinding approach. No regime provides an ordinary work-specific query linked to complete lifecycle records.

4. A Three-Tier Selective Verifiability Framework

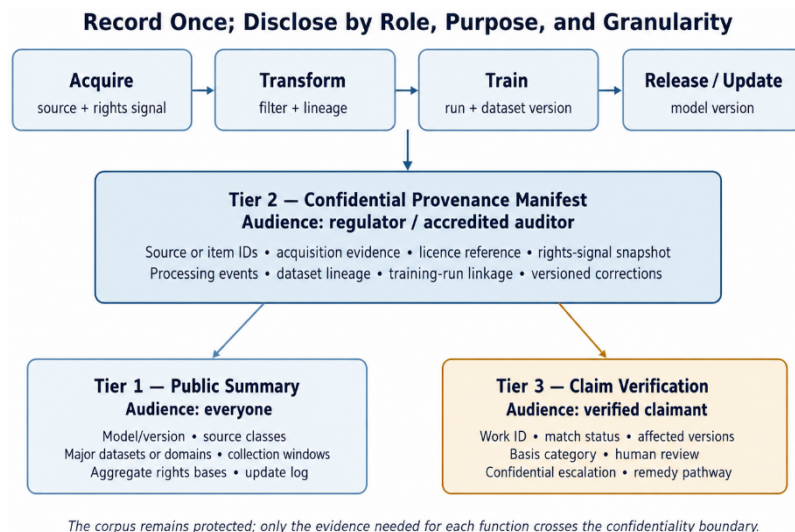
4.1. Design principles

The framework rests on six principles derived from proportionality, purpose limitation, procedural fairness, trade-secret protection, effective rights enforcement, and interoperable recordkeeping. Evidence before exposure means identifying the claim to be tested before deciding which information must be disclosed. Lifecycle linkage connects acquisition, transformation, dataset version, and model version. Data minimization limits each audience to information necessary for its function. Contestability prevents unreviewable assertions about legality, secrecy, or record completeness. Proportionality adjusts detail, audit frequency, and support to model scale, deployment, and provider capacity. Interoperability favors open schemas and exportable records over a proprietary government database. Together these principles explain why indiscriminate public disclosure, regulator-only access, ordinary litigation discovery, or cryptographic commitment alone is insufficient.

The architecture does not place all trust in one actor. Providers create and preserve records. An accredited auditor can test routine controls and report exceptions. A public authority can compel access, review contested trade-secret designations, and order correction. Rights holders initiate work-specific claims and can challenge an indeterminate or adverse response. Courts retain authority over infringement, fair use, license interpretation, and remedies.

Figure 1 maps the three access routes onto the training-data lifecycle.

<Figure 1 Selective verifiability across the training-data lifecycle>



4.2. Tier 1: public training-content summary

Tier 1 gives the public and rights holders enough information to orient further

inquiry. It should identify the provider, model and version, training stages covered, modalities, broad volume ranges, source categories, named major public datasets, crawler identities and collection windows, significant domains or source groups, aggregate rights-basis categories, treatment of rights reservations, key exclusion practices, use of synthetic or user data, and update dates. The EU template and AB 2013 provide workable starting fields.²⁹⁾

The summary should be standardized and machine-readable as well as human-readable. Consistent categories allow comparison across versions and providers. A change log should distinguish a new model from fine-tuning or continued training. Providers should identify known information gaps and explain why retrieval would be disproportionate, rather than silently omitting fields.

Tier 1 should not require a public list of every URL, work, license price, contract counterparty, weighting rule, or filtering threshold. Those fields can reveal protected compilation choices and are not necessary for broad orientation. Public disclosure should nevertheless be specific enough to generate a testable hypothesis. “A mixture of public and licensed data” is insufficient; “English-language web text collected by crawler C between specified dates, named datasets D1-D4, licensed news archives, and synthetic instruction data” is materially more useful without exposing an entire corpus.

4.3. Tier 2: confidential provenance manifest

Tier 2 is a structured provenance manifest maintained at source- or item-level granularity and available to a regulator or accredited auditor under statutory confidentiality. Without duplicating underlying content, it should record a stable identifier; source or acquisition channel and time; a fingerprint where appropriate; counterparty and recorded rights basis; license reference or rights-reservation snapshot; preprocessing and exclusion events; and lineage to dataset and model versions. Versioned entries allow sampling of whether public categories map to actual sources and agreements and whether exclusions were applied.

Granularity should reflect the source: an object may fit a web crawl, while a disclosed dataset release may be the feasible unit for licensed or legacy collections. Audit reports must state the population, sampling method, exceptions, and assurance level; sampling cannot certify corpus-wide legality or completeness. Secure review should use logging, download restrictions, and contractual segregation. A secrecy designation must identify likely harm and why narrower disclosure is inadequate and remain reviewable by the authorized reviewer.³⁰⁾

29) European Commission, *supra* note 19; Cal. Civ. Code § 3111(a).

30) Perry Keller & Tanya Aplin, *supra* note 5, pp. 175-184.

4.4. Tier 3: claim-triggered verification

Tier 3 permits a verified rights holder to test a specific work using an ISBN, DOI, canonical URL, platform identifier, hash, or robust fingerprint together with authority to act. The response identifies the covered model and dataset versions and returns a reasoned status: inclusion, exclusion, no match, or indeterminate. Inclusion may disclose the acquisition period, source and recorded legal-basis categories, and review route, but not unrelated works or full contracts; exclusion is limited to documented versions, no match is not proof of absence, and indeterminate requires human review.

Both parties may challenge the result with additional identifiers or objections to authority and legal basis. Authentication, rate limits, and bounded protective access may deter abuse. Disputed evidence should be preserved, and a court or regulator may authorize deeper confidential review. The channel neither displaces ordinary proceedings nor becomes a mandatory exhaustion requirement.

4.5. Legal safeguards, proof allocation, and data protection

Selective access concentrates sensitive information and therefore requires a lawful basis, purpose limitation, data minimization, role-based access, security logging, retention controls, and appropriate pseudonymization under the Personal Information Protection Act.³¹⁾ A manifest should ordinarily retain identifiers and acquisition evidence rather than copies of works; hashes are not necessarily anonymous. Cross-border access and content inspection should be limited to defined verification or enforcement purposes.

Proof allocation should track control of information. The claimant initially shows authority and identifies the work and model; the provider then reasonably searches relevant versioned records and gives a reasoned status. Each status describes records, not liability: inclusion does not establish infringement, and exclusion or no match does not prove corpus-wide absence. Validated integrity controls support only rebuttable inferences about a record or control. Falsification, destruction, or obstruction should have consequences only under express law and due process. The channel supplements ordinary remedies.

4.6. Technical implementation and evidentiary limits

Implementation can reuse W3C PROV and AI or dataset bills of materials such as SPDX and CycloneDX to link sources, transformations, dataset releases, licenses, and training runs. Digital signatures and append-only or Merkle commitments can make retrospective alteration harder and permit inclusion proofs without exposing the whole manifest. Legislation should prescribe outcomes—versioned provenance, integrity, exportability, auditable rights-basis records, and bounded responses—

31) Personal Information Protection Act (Republic of Korea), Arts. 3, (1), (2), (7).

while standards define formats.³²⁾

These techniques have strict limits. Exact hashes miss transformed content; robust fingerprints introduce false matches and misses. Inclusion proofs do not establish manifest completeness or lawful use, and authentic records may be false, so independent sampling remains necessary. Rights-reservation signals may change or conflict across terms, robots.txt, metadata, and platforms; the manifest should preserve the observed signal, location, time, and scope without deciding legal effect. Law should remain technologically neutral and treat cryptographic methods as evidentiary supports, not substitutes for review.³³⁾

5. A Policy Path for Korea

5.1. Relationship to Article 35-5 and substantive legality

Selective verifiability should be framed as a copyright-evidence and recordkeeping mechanism, not as a new copyright exception, an interpretation that enlarges Article 35-5, or a defense to duties under the AI Basic Act. The flexibility of the fair-use provision makes reliable facts especially important. Acquisition channels, access restrictions, license references, and contemporaneous rights signals may illuminate the circumstances and character of acquisition and use. Dataset lineage and version linkage may help identify the material and amount used, while output controls may be relevant to potential market substitution or prejudice. Provenance records cannot independently determine copyright ownership, license validity, the legal effect of a rights signal, substantiality, market harm, or the ultimate fair-use conclusion. Conversely, a fair-use conclusion does not displace an independently triggered notice, labeling, posting, explanation, or reporting obligation concerning copyrighted or non-copyright information.

The tiers perform different evidentiary functions in that inquiry. Tier 1 enables orientation and claim formation but cannot establish fair use or infringement. Tier 2 permits a competent reviewer to test whether public assertions correspond to contemporaneous acquisition, rights-basis, exclusion, and lineage records. Tier 3 communicates only the bounded result necessary for work-specific contestation. The progressive increase in access responds to evidentiary need, not to the presumption that the presence of copyrighted material is unlawful.

5.2. Legal form and institutional allocation

Korea should proceed through a staged allocation of responsibilities rather than

32) The Linux Foundation and its Contributors, “The System Package Data Exchange® (SPDX®) Specification Version 3.0.1”, SPDX, <<https://spdx.github.io/spdx-spec/v3.0.1/>>, 검색일: 2026. 7. 1.; CycloneDX, “Machine Learning Bill of Materials (ML-BOM)”, CycloneDX, <<https://cyclonedx.org/capabilities/mlbom/>>, 검색일: 2026. 7. 1.

33) Shayne Longpre et al., “Consent in Crisis: The Rapid Decline of the AI Data Commons”, Advances in Neural Information Processing Systems 37, 2024, pp. 1-46.

assign every function to a single institution. An initial pilot could be led by the Ministry of Culture, Sports and Tourism and the Korea Copyright Commission, in coordination with the Ministry of Science and ICT and, where personal information is implicated, the Personal Information Protection Commission. It should test the public-summary and confidential-manifest profiles, secure review, claimant matching, and auditor criteria with developers, collective management organizations, creator groups, open-model communities, and small providers. Its purpose is to generate evidence on cost, matching accuracy, security, and institutional capacity before mandatory duties are fixed.

A subsequent statutory scheme should locate the recordkeeping and claim-response duty in copyright legislation rather than derive it expansively from Articles 31 or 34 of the AI Basic Act. Article 31 presently addresses notice of AI-based services and identification of AI-generated outputs. Article 34 requires specified measures for high-impact AI, including a technically feasible explanation plan concerning an overview of training data and posting of the plan's principal content, but it does not supply a copyright-specific power to compel source-level training records or adjudicate secrecy claims. Copyright legislation should define covered providers and material model updates, specify required records and responses, authorize confidential access, and establish sanctions and review, while leaving Article 35-5's substantive standard unchanged.

Institutional roles should be separated by legal function. The Ministry of Culture, Sports and Tourism should hold implementing, compulsory-information, and administrative-sanction authority. The Korea Copyright Commission could administer technical profiles, operate or supervise the claimant gateway, maintain an auditor registry, support dispute resolution, and publish aggregate findings. The Ministry of Science and ICT should coordinate the scheme with the AI Basic Act, while the Personal Information Protection Commission retains authority over personal information. Independent auditors may provide routine assurance but should not make binding determinations of ownership, fair use, or infringement. The Commission's existing functions do not themselves create general power to compel private training records, accredit regulatory auditors, or impose penalties; those powers require express legislation.

Confidentiality safeguards should bind every official, auditor, expert, and contractor who receives protected information. Review should occur through a secure environment with access logs, export restrictions, field-level segregation, and sanctions for misuse. A provider claiming trade secrecy must identify the protected category, likely competitive harm, and why a less restrictive response is inadequate. The provider should not be the final judge: the competent authority or court should apply necessity and proportionality and permit review of an access or disclosure decision. Public reports should disclose methods, aggregate findings, and corrective actions without revealing protected substance.³⁴⁾

34) Act on Prevention of Unfair Competition and Protection of Trade Secrets, *supra* note 15, Art.

5.3. Proportional obligations and incentives

Coverage should turn on market placement and functional control. A provider that places a general-purpose generative model or service on the Korean market and controls a material training or fine-tuning stage should document the content used for that stage. A downstream modifier should document its additional data and identify the upstream model and public summary. A distributor with no access to or control over training records should not be required to recreate them but should transmit the upstream documentation and identify the responsible provider.

Proportionality should adjust how, not whether, minimum provenance is kept. Small providers and non-commercial researchers could use simplified templates, longer transition periods, open-source tools, sampled audits, and Commission support. Public summaries may use broader volume bands, and routine external audit frequency can reflect model scale and deployment risk. Open release should not create a blanket exemption from minimum source and rights-basis records: openness of weights does not answer how training content was acquired. Truly personal experiments, security-only systems, and defense systems may require exclusions or separate regimes.

Within that boundary, good-faith documentation may justify limited evidentiary and administrative incentives. Contemporaneous records, independently validated integrity controls, prompt correction, and cooperation with a verified query may support an inference about facts actually documented or mitigate a penalty for a recordkeeping violation. They should not support an inference that the training was lawful, create a safe harbor, or shift to the claimant the burden of disproving fair use. Deliberate destruction, material misrepresentation, or obstruction may justify escalating audit, an adverse evidentiary consequence, or an administrative sanction only where legislation expressly authorizes it. Failure to honor a rights reservation should be sanctionable under this framework only where the reservation is legally effective, contractually binding, or covered by an enforceable provider representation; the recordkeeping rule should not create a substantive opt-out by implication.³⁵⁾

A claim response should offer more than a binary answer. Available outcomes can include correction of the manifest, prospective exclusion from the next training cycle, preservation of disputed records, referral to licensing or collective-management channels, output testing and mitigation, and, where legally justified, compensation or judicial relief. Because complete “unlearning” may be technically difficult and may not remove all influence, the framework should not promise deletion as the universal remedy.

2(2): X.AI LLC v. Bonta, *supra* note 9.

35) Compliance with recordkeeping should not alter copyright-infringement elements or the substantive Article 35-5 inquiry.

5.4. Phased implementation

Table 2 presents an illustrative phased implementation sequence.

<Table 2 Proposed Korean implementation roadmap>

Phase	Period after adoption	Core actions	Assurance level
1. Standard and pilot	0–12 months	Publish Tier 1 and Tier 2 profiles; pilot claimant matching; test secure review; create auditor criteria; provide SME tools	Voluntary participation with public reporting of lessons
2. Baseline duty	12–24 months	Require public summaries and versioned internal manifests for newly released covered models; preserve source and rights-signal records	Self-attestation plus risk-based sampling
3. Verification channel	18–30 months	Launch verified claimant portal; require reasoned statuses and human review; authorize confidential escalation	Accredited audit and Commission supervision
4. Mature enforcement	30 months onward	Periodic audits for larger providers; penalties for material misrepresentation or obstruction; publish aggregate findings; revise technical profiles	Risk-based enforcement with appeal and judicial review

Records should be retained for the commercial life of a covered model and for a defined period, provisionally five years, after withdrawal or the last material update, subject to litigation holds and data-minimization rules. The period should be tested during the pilot against storage burden, model update cycles, and copyright limitation periods. Underlying copyrighted content need not be duplicated merely to satisfy the manifest duty where identifiers and acquisition evidence are sufficient.

The framework can also reinforce initiatives already identified in Korea's 2026 guide, including better rights-information infrastructure and licensing facilitation. A rights-information database is more useful when a provider can record which version of a rights signal or license entry it consulted. Conversely, provenance records can reveal recurring categories where transaction costs remain high and collective licensing or standard licenses may be warranted. Selective verifiability is therefore not a substitute for licensing markets; it supplies the information infrastructure on which more credible markets and adjudication can operate.

6. Conclusion

Existing law fragments transparency by function: Korea emphasizes user notice, output identification, and limited high-impact explanation; the EU combines operational, synthetic-content, and GPAI documentation duties; California separately addresses dataset summaries, content provenance, and frontier risk; and Japan relies chiefly on guidance. None creates a standard work-specific route from a

public training summary to lifecycle-linked evidence. Copyright fair use and TDM rules remain separate and do not excuse transparency duties covering protected and unprotected information.

Selective verifiability fills that procedural gap through a public summary, a confidential provenance manifest, and claimant-triggered verification. It does not determine training legality or guarantee record completeness; it makes relevant facts durable and secrecy claims reviewable while limiting unnecessary exposure. Korea should test the design through a phased pilot, proportionate recordkeeping, accredited audit, and an express copyright-law verification channel that preserves Article 35-5 and trade-secret protection. Future work should evaluate matching error, provider cost, privacy, cross-border access, auditor independence, and institutional capacity.

참고문헌

학술지(국내 및 동양)

- Insu Yang & Sunha Park, “A Study on Requirement of Non-Publicity for Trade Secrets”, *Journal of Intellectual Property*, Vol.20 No.1(2025).
- Jeonghun Park, “A Study on the Interpretation and Legislation of Data-Related Laws”, *Journal of Intellectual Property*, Vol.17 No.2(2022).
- Seongwon Kim et al., “An Exploratory Study on New Provisions for Exemption from the Use of Copyrighted Works in Text and Data Mining Processes—Applying to Generative Adversarial Networks”, *Journal of Intellectual Property*, Vol.18 No.1(2023).

학술지(서양)

- Adam Buick, “Copyright and AI Training Data—Transparency to the Rescue?”, *Journal of Intellectual Property Law & Practice*, Vol.20 No.3(2025).
- Mark A. Lemley & Bryan Casey, “Fair Learning”, *Texas Law Review*, Vol.99 No.4(2021).
- Pamela Samuelson, “Generative AI Meets Copyright”, *Science*, Vol.381 No.6654(2023).
- Perry Keller & Tanya Aplin, “Reconciling Trade Secrets and AI Public Transparency”, *The Cambridge Law Journal*, Vol.85 No.1(2026).
- Peter Henderson et al., “Foundation Models and Fair Use”, *Journal of Machine Learning Research*, Vol.24 No.400(2023).
- Shayne Longpre et al., “A Large-Scale Audit of Dataset Licensing and Attribution in AI”, *Nature Machine Intelligence*, Vol.6(2024).
- Timnit Gebru et al., “Datasheets for Datasets”, *Communications of the ACM*, Vol.64 No.12(2021).

판례

- X.AI LLC v. Bonta, No. CV 25-12295-JGB (SSCx), Order Denying Preliminary Injunction (C.D. Cal. 4 March 2026), appeal docketed, No. 26-1591 (9th Cir. 17 March 2026).

인터넷 자료

- Alexander Wan et al., “The 2025 Foundation Model Transparency Index”, arXiv, <<https://arxiv.org/abs/2512.10169>>, 작성일: 2025. 12. 11.
- CycloneDX, “Machine Learning Bill of Materials (ML-BOM)”, CycloneDX, <<https://cyclonedx.org/capabilities/mlbom/>>, 검색일: 2026. 7. 1.
- European Commission, “Explanatory Notice and Template for the Public Summary of Training Content for general-purpose AI models”, European Commission, <<https://digital-strategy.ec.europa.eu/en/library/explanatory-notice-and-template-public-summary-training-content-general-purpose-ai-models>>, 검색일: 2026. 7. 1.
- European Commission, “Transparency Obligations under Article 50 of the AI Act”, European Commission, <<https://digital-strategy.ec.europa.eu/en/faqs/transparency-obligations-under-article-50-ai-act>>, 검색일: 2026. 7. 1.
- Stanford Center for Research on Foundation Models, “Foundation Model Transparency Index”, Stanford Center for Research on Foundation Models, <<https://crfm.stanford.edu/fmti/December-2025/index.html>>, 검색일: 2026. 7. 1.
- The Linux Foundation and its Contributors, “The System Package Data Exchange® (SPDX®) Specification Version 3.0.1”, SPDX, <<https://spdx.github.io/spdx-spec/v3.0.1/>>, 검색일: 2026. 7. 1.
- World Wide Web Consortium, “PROV-DM: The PROV Data Model”, W3C, <<https://www.w3.org/TR/>>

prov-dm/>, 작성일: 2013. 4. 30.

기타 자료

- Agency for Cultural Affairs (Japan), “General Understanding on AI and Copyright in Japan—Overview”, Agency for Cultural Affairs, 2024.
- AI Strategy Headquarters (Japan), “Guidelines on Ensuring Appropriateness in the Research, Development and Utilization of Artificial Intelligence-related Technology”, AI Strategy Headquarters, 2025.
- Ministry of Culture, Sports and Tourism & Korea Copyright Commission, “Guide on Applicability of the Fair Use Doctrine to Training of Generative AI Models”, Korea Copyright Commission, 2026.
- Ministry of Science and ICT (Republic of Korea), “Guidelines for Securing AI Transparency”, Ministry of Science and ICT, 2026.
- Shayne Longpre et al., “Consent in Crisis: The Rapid Decline of the AI Data Commons”, *Advances in Neural Information Processing Systems* 37, 2024.
- United States Copyright Office, “Copyright and Artificial Intelligence, Part 3: Generative AI Training”, United States Copyright Office, 2025.
- World Intellectual Property Organization, “WIPO Guide to Trade Secrets and Innovation”, WIPO, 2024.

생성형 인공지능(AI) 도구 활용

- Anthropic. (2026), Claude, [대형언어모델(LLM)], <https://claude.ai/>, 원고의 언어 표현 및 가독성 개선에 활용하였음. 저자는 모든 생성 결과물을 검토·수정하였으며, 출판물의 내용에 대한 최종 책임을 짐.