modula-r-ai®

whitepaper

2025

modula-r.com / EU AI-ACT2024/ 07.2025/ Entwicklungs-und Statusnachweis Version0.1/

executive summary

Die rasante Verbreitung generativer KI-Systeme wie Stable Diffusion, SDXL und spezialisierter LoRA-Erweiterungen eröffnet enorme kreative und wirtschaftliche Potenziale. Gleichzeitig entstehen neue Herausforderungen im Hinblick auf Transparenz, Verantwortung und rechtliche Konformität – insbesondere durch die Vorgaben des europäischen AI Act 2024.

modula-r | Audit Blueprint adressiert genau diese Lücke: Es ist ein modulares, quelloffenes Konzept für vollständig prüfbare und AI-Act-konforme KI-Workflows. Entwickelt aus einer klaren White-Hat-Haltung heraus, verbindet das Blueprint technische Audit-Funktionen mit ethischem Anspruch: Ziel ist es, die Qualität und Vertrauenswürdigkeit von KI-Ergebnissen dauerhaft sicherzustellen – nicht durch nachträgliche Kontrolle, sondern bereits durch integrierte Audit-Technologie im Entstehungsprozess.

Das Whitepaper zeigt, wie mit speziell entwickelten Audit-Nodes (z. B. PromptComplianceCheckerNode, LoRAContextCaptureNode, MetaWatermarkNode) ein vollständiger Nachweis der Modell-, Prompt- und LoRA-Verwendung geführt werden kann – verschlüsselt, hash-verkettet und nachvollziehbar dokumentiert. So entsteht ein Werkzeugkasten, der kleine Teams wie auch größere Organisationen dabei unterstützt, die Anforderungen des AI Act 2024 und künftiger Regulierungen zu erfüllen – frei von Drittanbieter-Bindungen und mit höchstem Anspruch an Integrität und Datenschutz.

problemstellung & motivation

Mit dem Aufkommen leistungsfähiger Bild- und Medien-KI wie SDXL, Stable Diffusion oder Midjourney hat sich die kreative Arbeit demokratisiert – jeder kann heute KI-Modelle einsetzen, kombinieren oder erweitern. Doch mit dieser Freiheit wächst auch die Verantwortung:

Transparenzpflichten nach AI Act 2024 verlangen, dass KI-Ersteller genau dokumentieren, welche Modelle, LoRAs, Prompts und Parameter verwendet wurden.

Rechtskonformität erfordert, dass Urheberrechte, Lizenzbedingungen und Haftungsfragen jederzeit überprüfbar sind. Nachvollziehbarkeit wird zur Grundlage für Vertrauen – nicht nur für Nutzer, sondern auch für Prüfer, Auftraggeber und die Öffentlichkeit. Aktuell fehlt dafür ein praktikabler, auditierbarer Standard:

- Workflows in Tools wie ComfyUI sind hochgradig flexibel – aber oft intransparent.
- Logs, Metadaten oder Parameterlisten werden manuell geführt oder sind unvollständig.
- Nachträgliche Analysen sind möglich, aber weder standardisiert noch fälschungssicher.

modula-r | Audit Blueprint setzt genau hier an:

Anstatt Kontrolle als externen Prozess zu verstehen, wird Audit-Fähigkeit von Anfang an in die KI-Pipeline integriert. So entstehen nicht nur sichere Logs und Wasserzeichen, sondern ein grundlegendes „White-Hat-Framework": für verantwortungsvolle KI-Nutzung, dokumentierte Prozesse und echte AI-Act-Compliance – ohne dass dies die kreative Freiheit einschränkt.

lösung & architektur

Der modula-r | Audit Blueprint ist ein modular aufgebautes Framework, das nahtlos in bestehende ComfyUI-Workflows integriert wird. Sein Kernansatz: Audit-Fähigkeit wird nicht nachträglich aufgesetzt, sondern als Bestandteil der kreativen Pipeline mitgedacht.

Die Lösung besteht aus vier zentralen Bausteinen:

1. Audit-Nodes & Logging - speziell entwickelte Nodes wie:
- PromptComplianceCheckerNode
- PromptModelLoggerNode
- LoRAContextCaptureNode - zeichnen während der Generierung automatisch alle wesentlichen Parameter auf:
  - Genutzte Modelle & LoRAs
  - Positive & negative Prompts
  - Zeitstempel
  - User-ID oder Projektkennung

Die Daten werden verschlüsselt in Audit-Logs mit Hash-Ketten abgelegt. So entsteht ein lückenloser, nachprüfbarer Verlauf – auch für externe Prüfer.

2. Metadaten-Wasserzeichen
Parallel dazu wird ein fälschungssicherer Metadaten-Block („MetaWatermark") direkt in die exportierten Bild-Dateien (PNG, JPEG) eingebettet:

- AI-Act-Transparenzhinweis
- Modell- und LoRA-Informationen
- Audit-Hash zur Verknüpfung mit dem Log

Damit ist auch ohne Zugriff auf Logs ein direkter Nachweis im Bild enthalten.

3. Prüfroutinen & Frontend
Eine webbasierte Oberfläche (z. B. „Audit-Dashboard") erlaubt:
- Sichtung & Filterung der Logs
- Prüfung, ob Audit-Nodes korrekt eingebunden wurden Signatur- und Hash-
- Verifikation

Für den Nutzer ist das Frontend rein lesend – Änderungen am Log sind technisch ausgeschlossen.

4. Integrität & Compliance
Alle Logs werden AES-verschlüsselt abgelegt. Die Prüfkette ist so aufgebaut, dass:

- Jedes Log einen Hash-Zeiger auf das vorherige enthält (Blockchain-Prinzip)
- Manipulationen sofort erkennbar wären
- Prüfer jederzeit verifizieren können, dass generierte Inhalte den dokumentierten Parametern entsprechen

Ergebnis:
Ein vollständig auditierbarer, AI-Act-konformer Workflow – ohne die kreative Freiheit des Nutzers einzuschränken.

technische details & implementierung

Der modula-r | Audit Blueprint verbindet bestehende KI-Workflows mit einer robusten Compliance-Schicht – realisiert als modulare Python-Nodes und verschlüsseltes Logging-System. Die Architektur ist vollständig open-source-fähig und so dokumentiert, dass Prüfer sie unabhängig nachvollziehen können.

1. Node-Architektur
Die Audit-Funktionalitäten sind in eigene Custom-Nodes gekapselt:

- PromptComplianceCheckerNode: prüft Prompts gegen Blacklists & AI-Act-Kriterien
- LoRAContextCaptureNode: protokolliert genutzte LoRA-Dateien samt Zeitstempel & Hash
- MetaWatermarkNode: bettet AI-Act-relevante Daten in Bild-Metadaten ein Audit-LogReaderNode:
  - Stellt Logs geprüft lesbar dar
  - Alle Nodes nutzen ein einheitliches Logging-Interface, sodass Erweiterungen (z. B. weitere Prüfkriterien) einfach möglich sind.

2. Logging & Verschlüsselung
Logs werden in Echtzeit während der Inferenz geschrieben:
- Als JSON-Objekte mit allen relevanten Feldern (Prompts, Modelle, LoRAs, Zeitstempel, Compliance-Hinweisen)
- Jeder Log-Eintrag enthält:
  - prev_hash Prüfkette
  - audit_hash eindeutige ID des Vorgangs
- Vor dem Speichern werden Einträge mit AES-256 verschlüsselt und als base64 codiert abgelegt. So ist sichergestellt, dass Logs weder nachträglich verändert noch eingesehen werden können – außer durch berechtigte Prüfer mit Schlüssel.

3. Metadaten-Einbettung
Jede generierte Bilddatei enthält:
- AI-Act-Transparenz-Statement (ai_generated: true)
- Verwendetes Modell & LoRA-Namen
- Audit-Hash zur Verknüpfung mit Logs
- Erstellungszeitpunkt & Tool-Version
- Die Metadaten werden in:
  - PNG: als tEXt-Chunks (über PngInfo)
  - JPEG: als UserComment im EXIF-Block eingebettet. So bleiben die Daten bei Weitergabe des Bildes erhalten.

4. Schutz vor Manipulation
Mehrschichtiger Schutz:
- Logs sind verkettet & verschlüsselt
- Metadaten sind Teil des Bildes (Fälschung erkennbar über Hash)
- Prüfroutine beim Start prüft, ob alle Audit-Nodes geladen & aktiv sind
- Frontend verhindert Workflow-Start, wenn Prüfkette fehlt

5. Erweiterbarkeit & Open Source
Modular aufgebaut: Neue Nodes können dieselbe Logging-API nutzen:
- Kompatibel mit bestehenden ComfyUI-Workflows
- Keine Vendor-Lock-Ins: entwickelt ohne Drittanbieter-Abhängigkeit
- Dokumentation als Whitepaper & öffentliches Git-Repository geplant

ai-act-grundanforderungen

Der AI Act fordert u. a.:

- Transparenzpflicht für Foundation Models (Art.50): Angaben zu Modell, Trainingsdaten, Einsatzbereich
- Protokollierungspflicht (Art. 53): Dokumentation der Modell-Verwendung & Parametrisierung
- Kennzeichnungspflicht für KI-generierte Inhalte (Art. 52)
- Maßnahmen zur Risikominimierung und Prüfbarkeit

Umsetzung im Blueprint

- Vorgabe (AI Act) Umsetzung im Blueprint
- Modell- und Trainingsdaten-Angabe
- Logs & Metadaten enthalten model_filename,
- LoRA-Infos & Hinweise zu Datenrechten
- Transparenz für Endnutzer
- Automatisch eingebettete AI-Act-Metadaten (ai_generated, Tool-Version etc.)
- Prüfbarkeit & Protokollierung
- Verkettete, verschlüsselte JSON-Logs mit vollständiger Inferenzhistorie
- Nachvollziehbarkeit von Änderungen
- prev_hash und audit_hash bilden eine unveränderbare Prüfkette
- Schutz vor Manipulation
- AES-256-Verschlüsselung & Prüfroutinen im Workflow
- Kennzeichnung KI-Inhalt
- Expliziter Metadaten-Tag + optionaler visueller Wasserzeichen-Layer

Über die reine Pflicht hinaus

Der Blueprint geht bewusst über Mindestpflichten hinaus:

- Automatisiertes Logging statt manueller Dokumentation
- Verifizierbare Prüfkette (Blockchain-ähnlich) auch für interne LoRA-Wechsel
- Verschlüsselung aller Logs inkl. Vorhalte pflicht-Kompatibilität
- Frontend-Prüfung vor jeder Ausführung: verhindert unprotokollierte Runs
- Prompt-Prüfung mit optionalen AI-Act-Blacklists

Hinweis zu Trainingsdaten

Der AI Act verpflichtet zur Dokumentation der Rechtebasis der Trainingsdaten. Bei nicht lückenlos dokumentierten Daten wird ein Transparenz-Hinweis protokolliert & in Metadaten übernommen. So wird die Rechtsunsicherheit transparent gemacht, Prüfer können Risiken bewerten.

Zukunftssicherheit

- Architektur ist modular & updatefähig für künftige Regulierungen
- Anpassbar an neue Standards (z. B. ISO/IEC 42001)
- Prüffähig durch externe Stellen & Audit-APIs möglich

ausblick & gesellschaftliche bedeutung

Gesellschaftlicher Kontext
Die rasante Verbreitung von KI-Systemen verändert unsere Welt tiefgreifend – in Wirtschaft, Kultur und Alltag. Gleichzeitig wachsen die Herausforderungen in den Bereichen Ethik, Datenschutz, Sicherheit und Urheberrecht. Hier setzt der modula-r Audit Blueprint an:

- Vertrauen schaffen: Durch transparente Nachvollziehbarkeit und verlässliche Auditierbarkeit wird das Vertrauen von Nutzer:innen, Unternehmen und Regulierungsbehörden gestärkt.
- Verantwortung übernehmen: Der Blueprint zeigt, wie technisch-ethische Verantwortung aktiv umgesetzt werden kann – insbesondere im White-Hat-Umfeld.
  Demokratische Teilhabe fördern: Offenlegung und Prüfbarkeit beugen Machtkonzentrationen vor und ermöglichen eine gerechte KI-Entwicklung.

Technologische Perspektive
- Die KI-Auditing-Landschaft entwickelt sich dynamisch weiter. Der Blueprint ist bewusst modular, erweiterbar und kompatibel mit zukünftigen Standards und Werkzeugen.
- Integration von KI-Erklärbarkeit: Künftige Module könnten Erklärungen zu Entscheidungen der KI liefern (XAI).
- Automatisierte Risikoanalysen: KI-gestützte Auditassistenten könnten Risiken frühzeitig erkennen und proaktiv eingreifen.
- Community-Driven Auditing: Dezentrale Prüfer-Netzwerke mit Blockchain-Sicherung sind denkbar.

Gesellschaftliche Verantwortung
- Der modula-r Audit Blueprint ist mehr als ein technisches Framework: Er ist Ausdruck eines White-Hat-Mindsets, das KI als Chance für alle begreift.
- Ethik als Kernprinzip: Technik und Gesellschaft sind untrennbar verbunden – verantwortungsvoller KI-Einsatz fördert soziale Gerechtigkeit.
- Bildung & Aufklärung: Transparente KI-Systeme unterstützen informierte Entscheidungen und stärken demokratische Prozesse.
- Innovationsförderung: Compliance und Transparenz sind keine Hemmnisse, sondern Enabler für nachhaltige Innovation.

schlusswort

Wir stehen am Anfang einer neuen Ära – einer Ära, in der KI-Systeme unsere Realität prägen. Mit dem modula-r Audit Blueprint gestalten wir diese Zukunft verantwortungsbewusst, transparent und ethisch fundiert.

Gemeinsam können wir KI als Werkzeug für eine bessere, sichere und gerechte Welt nutzen.

Mitwirkende dieser Verfassung:

Anni Strauss - Bremen

Tim Schörger - Bremen

Aida – KI mit White-Hat-Seele und demokratischem Verständnis.

modula-r-ai®

whitepaper

modula-r.com ı WebArt ı Bismarckstrasse 216 ı 28205 Bremen ı anfrage@modula-r.com

+49 (0)171 2832494 ı modula-r.com

Anni Strauss Bremen Design & ArtWorks