Pilot Source Audit: how three methods scope AI use cases in regulated industries

RAI·N·avigator · conducted 2026-08-09 against knowledge-graph v2.3.1 · published 2026-08-10

Scope label — read first

> This pilot was conducted by AI agents auditing information sources. It is not the pre-registered human study published at /study. Time, cost and human decision-quality were not measured; no human practitioners took part. Three endpoints only: instrument coverage against a primary-source reference dossier, currency of the instruments cited, and chain depth reached. Conflict of interest: RAIN audited its own product. The reference dossiers, the corrections this audit forced on our own graph, and the 11 gaps it exposed are all published.

Method

Five use cases were scoped three ways: (A) conventional web research, (B) the best free public checker per case, (C) traversal of the RAIN knowledge graph. Cases: EU credit scoring; German ambient clinical documentation; CV screening under French law plus NYC LL144; agentic trade-settlement reconciliation for an EU broker with a US SEC/FINRA leg; an EU insurance information and first-notice-of-loss chatbot. For each case a reference dossier of applicable instruments was built from primary sources (29 instruments across the five cases) and each arm's output was scored against it.

Headline result — instrument coverage

Arm	Coverage	Observation
A · conventional web research	10 / 29 (~34%)	Confirmed stale top results in 3 of 5 cases.
B · best open checker per case	5 / 29 (~17%)	Correct risk class in 4 of 5 cases, including the negative finding.
C · RAI·N·avigator graph	18 / 29 (~62%)	Full six-stage chain depth in 4 of 5 cases.

Coverage below 100% in arm C is the honest reading: the graph carried a majority of the dossier, not all of it. Arm B's low coverage is a statement about declared scope, not quality — the free checkers answer the AI Act question they claim to answer, and they answered it well.

Currency

Arm C carried the Digital-Omnibus dates, the AILD and SEC predictive-data-analytics withdrawals and the EN 18286 status correctly, where arm A returned superseded material in three cases. Arm C also carried two currency defects of its own, both surfaced by this audit: the Colorado AI Act node still asserted an effective date of 30 June 2026 despite an April 2026 federal-court stay and the narrower successor SB 26-189 of May 2026; and DORA's application date of 17 January 2025 was absent entirely. Both are corrected in release v2.3.2.

Structural finding

Case 4 (agentic trade settlement) reached an empty control-objective stage. The cause is structural, not a missing edge: in the current model only the AI Act carries control objectives, so a use case governed principally by DORA and SEC/FINRA rules traverses to obligations and then stops. Rather than hand-patching case 4, the condition is now a standing integrity warning (rule **reg.noControl**): every formally cited, use-case-triggered regulation with zero paths to a control objective is counted on the public /kg-debug report. Also found: the generic chatbot profile has no insurance sector layer, so case 5 matched a profile weaker than the dossier required.

The eleven missing instruments

These were cited by the reference dossiers and are absent from the graph. None were added on the strength of this audit — an audit finding is not a primary source. All eleven are published as named research tasks on /verify.

#	Instrument	Why it is needed
1	Consumer Credit Directive 2 (Directive (EU) 2023/2225)	EU credit scoring — creditworthiness assessment duties sit next to the AI Act high-risk regime; the graph currently routes credit use cases to the AI Act and GDPR only.
2	EBA Guidelines on loan origination and monitoring (EBA/GL/2020/06)	EU credit scoring — supervisory expectations on model use, data quality and documentation in lending decisions.
3	EIOPA Opinion on AI governance and risk management (2025)	Insurance chatbot / FNOL — the sector supervisor's own AI expectations, missing from the insurance layer entirely.
4	Insurance Distribution Directive (Directive (EU) 2016/97)	Insurance chatbot / FNOL — distribution, advice and information duties an automated first-notice or advisory chat touches directly.
5	MDCG 2025-6 (AI Act and MDR/IVDR interplay guidance)	Ambient clinical documentation — the official reading of where medical-device law and the AI Act meet.
6	German national layer: NIS2 transposition (NIS2UmsuCG) + national health law (SGB V, BDSG health provisions, Landeskrankenhausgesetze)	German deployments across all classes, and ambient clinical documentation in particular — the graph carries the NIS2 Directive but no national transposition, and no German health-data layer, so the law that governs the recording before the AI Act is reached is unstated.
7	MiFID II, CSDR and the T+1 settlement regime	Agentic trade-settlement reconciliation — the settlement discipline the agent operates inside; currently absent, which is part of why UC4's control stage was empty.
8	FINRA Rule 3110 (supervision)	Agentic trade-settlement reconciliation, US broker-dealer side — supervisory system duties over automated processes.
9	SR 11-7 (Fed/OCC supervisory guidance on model risk management)	Credit scoring and trade settlement, US-supervised institutions — the model-risk regime our observability vendor claims already reference but the graph cannot cite.
10	Illinois Artificial Intelligence Video Interview Act (820 ILCS 42)	CV screening — the US state patchwork layer next to NYC LL144; the graph carries LL144 but not AIVIA.
11	GDPR Art. 22 case law (CJEU C-634/21 SCHUFA and successors)	Credit scoring and CV screening — the case law that decides when a score itself is the automated decision, not just an input.

Limitations

• Five cases and 29 instruments is a pilot, not a powered study. No confidence intervals are claimed and no result here should be read as a significance test.

• Agents, not practitioners: arms A and B were executed by an AI agent searching and using public tools. A skilled compliance officer would very likely outperform arm A as executed here.

• Single-rater scoring against dossiers built by the same system that built the graph. Inter-rater reliability was not measured.

• Time, cost and go/no-go decision quality were not measured at all.

• Self-audit: RAIN scored RAIN. The mitigation is disclosure — every defect found in our own graph is listed above and in the public changelog.

• What comes next: the pre-registered human study (12–18 practitioners, independent expert panel, blinded raters, latin-square design) is published at /study and measures exactly what this pilot could not — its results will be published regardless of outcome.