TAF: A Semantic Governance Framework for Assessing EU AI Act Article 15 Safeguards in AI-Based Digital Forensic Tool Procurement

Darren McCabe[1], David Lewis[1]

[1]ADAPT Centre, Trinity College Dublin, Ireland

Abstract

AI-based digital forensic tools are increasingly being considered for organisational and public-sector investigations into insider threats, data leakage, fraud, policy breaches, unauthorised access, and cybersecurity incidents. These tools may use machine learning or Large Language Models to classify artefacts, summarise communications, generate timelines, identify relevant evidence, and support investigative decision-making. In rights-sensitive digital forensic contexts, their outputs may affect employees, contractors, suspects, and other data subjects, and may influence disciplinary, employment, legal, or regulatory outcomes. This creates a procurement governance problem: buyers must assess whether AI-based forensic tools provide explicit safeguards for accuracy, robustness, cybersecurity, traceability, human oversight, and fundamental rights before deployment.

This paper proposes the Trustworthy AI Forensics (TAF) framework, a semantic governance framework for assessing whether procurement documentation for AI-based digital forensic tools operationalises safeguards derived from the EU Artificial Intelligence Act, with a particular focus on Article 15 accuracy, robustness, and cybersecurity requirements [1, 2]. TAF does not presume that every forensic AI tool is legally high-risk: applicability remains a separate human legal assessment based on intended purpose and deployment context. The framework represents legal provisions, legal sub-provisions, competency questions, governance safeguards, evidence indicators, risk dimensions, and structural procurement risks through the Mapping Trustworthy AI to Law Ontology (MTALO), paired with the deterministic Mapping Accuracy Quality Assessment method for detecting explicit safeguard evidence.

The prototype uses RDF/XML, SHACL, SPARQL, JSON taxonomy files, Python, and HTML reporting. Validation confirms end-to-end traceability from legal provisions to structural procurement-risk outputs. Three retained public-procurement cases demonstrate documentary evidence coverage, advisory gaps, readiness, and structural procurement risk across distinct AI value-chain configurations. The contribution is an end-to-end, machine-readable traceability model that links legal requirements to procurement questions, safeguards, evidence indicators, and reviewable risk outputs without claiming automated legal compliance.

Keywords

Trustworthy AI, Digital Forensics, EU AI Act, Article 15, AI Procurement, Semantic Web, Ontology, Governance Safeguards, Structural Procurement Risk, Fundamental Rights

1. Introduction

Digital forensic practice is increasingly shaped by artificial intelligence. Organisations now process large volumes of endpoint data, cloud logs, identity records, email exports, chat messages, collaboration-platform records, file metadata, and security alerts. AI-based tools promise to reduce investigation time by classifying artefacts, filtering large datasets, summarising communications, extracting entities, identifying relevant events, and assisting forensic analysts in constructing timelines and investigative narratives. Large Language Models further extend these possibilities through question answering over evidence collections, semantic search, and draft report generation.

These capabilities create opportunities, but they also introduce significant governance risks. Digital forensic investigations are not ordinary analytics exercises. They may produce or influence evidence used in disciplinary proceedings, employment decisions, litigation, regulatory reporting, fraud investigations, or referrals to law enforcement. In in-house investigations, affected individuals may include

NXDG 2026: NeXt-generation Data Governance workshop, September 15, 2026, co-located with 22nd SEMANTiCS, September 15–17, 2026, Ghent, Belgium

mccabed9@tcd.ie (D. McCabe); delewis@tcd.ie (D. Lewis)

https://www.tcd.ie/ (D. McCabe); https://www.tcd.ie/ (D. Lewis)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-13

CEUR-WS.org/Vol-4327/paper5.pdf

employees, contractors, customers, or third parties whose communications, system activity, documents, or metadata are examined. Errors in AI-supported forensic analysis can therefore affect privacy, data protection, non-discrimination, procedural fairness, effective remedy, and confidence in the reliability of digital forensic practice [3, 4].

The procurement stage is a critical control point in the responsible adoption of AI systems. If an AI-based forensic tool is acquired without explicit requirements for accuracy, robustness, cybersecurity, logging, data governance, human oversight, and safeguards for fundamental rights, those gaps may become embedded in investigative workflows. Procurement documentation often contains broad supplier claims such as "accurate", "secure", "compliant", "transparent", or "explainable". Such claims may be insufficient unless they are linked to measurable safeguards, documentary evidence, contractual obligations, validation data, and reviewable risk outputs.

The EU Artificial Intelligence Act imposes Article 15 accuracy, robustness, and cybersecurity requirements on high-risk AI systems [1, 2]. A digital forensic tool is not automatically a high-risk AI system merely because it supports an investigation; classification depends on its intended purpose, deployment context, and the Article 6 and Annex III criteria. TAF therefore treats legal applicability as a prior human-review question. Where Article 15 applies, its safeguards are binding requirements; where it does not, the same safeguards may still be adopted as procurement good practice. In either case, buyers need structured evidence rather than unsupported claims.

This paper proposes the Trustworthy AI Forensics framework, referred to as TAF. TAF is a semantic governance framework for procuring AI-based digital forensic tools, centred on the Mapping Trustworthy AI to Law Ontology (MTALO). It operationalises legal, technical, forensic, and procurement requirements into a machine-readable structure of governance safeguards, evidence indicators, risk dimensions, and structural procurement-risk outputs. In MTALO these are implemented as semantic model concepts rather than ordinary prose labels, with camel-case identifiers such as `GovernanceSafeguard` used only where the paper refers to ontology classes. The framework is supported by the Mapping Accuracy Quality Assessment method, referred to as MAQA, which performs deterministic screening of procurement documentation for explicit textual evidence of safeguards.

The contribution is an end-to-end, machine-readable traceability method linking legal requirements and applicability questions to competency questions, procurement safeguards, documentary evidence indicators, and structural procurement-risk outputs. TAF does not replace legal classification, model validation, forensic expert review, or conformity assessment.

The paper makes four contributions: a semantic traceability model for AI-based digital forensic procurement; an operational mapping of Article 15 safeguards to procurement evidence; a deterministic method for identifying explicit evidence in procurement artefacts; and a technical demonstration using ontology validation and three procurement value-chain cases.

2. Background and Related Work

Digital forensics involves the identification, preservation, collection, examination, analysis, and presentation of digital evidence. Traditional forensic practice depends on evidence integrity, chain of custody, repeatability, documentation, expert interpretation, and defensible methods [5, 6]. AI-based forensic tools can assist with artefact classification, anomaly detection, malware triage, image analysis, entity extraction, communication review, and timeline generation. However, such tools can also alter the evidential process by prioritising artefacts, summarising evidence, suggesting investigative hypotheses, or producing confidence scores that influence human judgement.

The risks are particularly significant in in-house investigations. These investigations may be conducted by corporate security, cyber resilience, insider threat, fraud, legal, HR, compliance, or internal audit teams. They may be concerned about suspected data exfiltration, unauthorised access, system misuse, intellectual property theft, policy breaches, harassment, fraud, malware incidents, or insider risk activity. Although organisations have legitimate interests in investigating misconduct and protecting systems, affected individuals retain rights relating to privacy, data protection, proportionality,

non-discrimination, fair procedures, and effective remedy [3, 4].

The EU AI Act establishes a risk-based framework and imposes requirements on systems classified as high-risk, including data governance, documentation, record-keeping, transparency, human oversight, accuracy, robustness, and cybersecurity [1, 2]. Before Article 15 safeguards are treated as binding requirements, the system's intended purpose and deployment context must first be considered under Article 6 and, where relevant, Annex III. Article 15 requires appropriate levels of accuracy, robustness, and cybersecurity throughout the lifecycle; related Article 13 transparency and instructions-for-use requirements and Article 14 human-oversight requirements provide supporting procurement context where they affect the evidence expected from providers. For digital forensic procurement, this creates both an applicability question and an evidence question: what task-specific validation, robustness testing, cybersecurity controls, limitations, and oversight information should protect models, prompts, logs, evidence data, outputs, and affected persons?

The EU Model Contractual Clauses for AI procurement provide a bridge between AI Act obligations and procurement practice by expressing legal and technical requirements as contract clauses and annexes [7]. However, procurement teams still require structured methods for checking whether supplier documentation contains explicit evidence of these requirements. Recent AI procurement standardisation is also relevant. IEEE 3119-2025 defines a procurement standard for artificial intelligence and automated decision systems, covering problem definition, solicitation preparation, vendor and solution evaluation, contract negotiation, contract monitoring, and AI-specific risk management across the procurement lifecycle [8]. TAF complements this state of the art by focusing on rights-sensitive digital forensic procurement and by representing evidence requests, review gaps, and risk outputs in a machine-readable semantic model.

Ontologies and semantic web technologies provide a promising foundation for such assessment. Standardised knowledge graph formats and graph processing languages including RDF, OWL, SPARQL, and SHACL enable machine-readable representation, querying, and validation of domain knowledge [9, 10, 11, 12]. In AI governance, ontologies can link legal provisions, technical controls, organisational responsibilities, evidence artefacts, and risk categories. For procurement, this enables a reusable vocabulary and traceability model that connects legal obligations to operational safeguards and documentary evidence.

3. Research Problem and Aim

The research problem is the lack of a structured, traceable, and machine-readable method that distinguishes legal applicability from documentary evidence assessment and then tests whether procurement documentation explicitly operationalises Article 15-related and fundamental-rights safeguards.

Current AI procurement assessments often depend on supplier claims, generic security questionnaires, informal expert judgement, or broad contractual statements. These approaches may be insufficient for AI-based forensic tools because they do not consistently connect legal obligations to specific evidence requirements. A supplier may state that a tool is accurate, robust, secure, or explainable without providing task-specific accuracy metrics, robustness testing, cybersecurity controls, audit logging, human oversight mechanisms, or evidence of its impact on rights.

This creates four linked gaps. The first is a legal-to-technical translation gap: Article 15 is expressed at a regulatory level, whereas procurement teams require operational requirements and supplier evidence. The second is a forensic reliability gap: digital forensic practice depends on evidence integrity, traceability, repeatability, and expert review [5, 6]. The third is a fundamental rights gap: in-house investigations may process sensitive personal data and influence serious employment or legal outcomes, but must do so while protecting the rights of individuals affected by those outcomes [3, 4]. The fourth is a standardisation gap: without a reusable semantic structure, different procurement teams may ask inconsistent questions and interpret safeguards inconsistently.

The research question is:

How can semantic web technologies support a human-reviewed applicability assessment

and a traceable evaluation of Article 15 and fundamental-rights safeguards in the procurement of AI-based digital forensic tools?

The aim is to develop and technically evaluate a semantic governance framework that records applicability questions, assesses procurement documentation against Article 15-related safeguards, and identifies related fundamental-rights and forensic-governance risks.

4. Methodology

The research follows a design-oriented methodology. The purpose is to construct and evaluate a governance artefact that prototypes semantic-driven procurement assessment. The framework is not designed to perform model benchmarking, legal certification, or automated compliance determination. Instead, it supports structured decision-making by making legal and forensic safeguards explicit and traceable within procurement documentation.

The design process has five stages. First, authoritative legal and governance sources are identified, including the EU AI Act as amended, EU Model Contractual Clauses for AI procurement, IEEE 3119-2025, public procurement principles, GDPR, the Charter of Fundamental Rights, and digital forensic principles [13, 4, 1, 2, 7, 8]. Second, source provisions are decomposed into granular sub-provisions and competency questions, then translated into procurement-level safeguards and evidence indicators. Third, these mappings are represented in the semantic model with links back to their sources. Fourth, deterministic screening evaluates procurement artefacts for explicit evidence. Fifth, coverage, traceability, risk summaries, and follow-up actions are reported. This is a document-analysis and artefact-design method rather than an authoritative doctrinal interpretation; ambiguous applicability and interpretation questions are reserved for human legal review.

The ontology development follows the Linked Open Terms methodology, which supports requirements specification, implementation, publication, and maintenance of ontology artefacts [14]. The ontology links legal provisions, granular sub-provisions, competency questions, governance safeguards, evidence indicators, risk dimensions, and structural procurement-risk outputs. In the RDF implementation these are semantic model concepts with explicit relation types; the paper uses camel-case names only where it refers to implementation identifiers.

The left-to-right arrows in Figure 1 are a reader-oriented traceability order rather than an assertion that each class is a `prov:wasDerivedFrom` derivation of the previous class. In the implementation, provenance is represented separately using PROV-O-style links from generated artefacts back to source material and script runs, for example `prov:wasDerivedFrom` for output provenance [15]. Domain links in MTALO are expressed as named traceability relations between requirements, questions, safeguards, evidence indicators, and risk outputs.

The prototype evaluation assesses whether the semantic links can be represented, validated, and used to produce reproducible procurement-governance outputs. It is a technical artefact and demonstration evaluation: it does not determine legal compliance, establish high-risk status, validate model performance, or test usability with practitioners.

The screening layer uses deterministic HIT/MISS evidence detection. A HIT means that explicit textual evidence was detected for a GovernanceSafeguard. A MISS means that expected evidence was not detected in the analysed procurement material. Coverage is calculated as:

$$\textit{Coverage} = \text{evidenced GovernanceSafeguards} / \text{expected GovernanceSafeguards}$$

This coverage score indicates alignment with documentary evidence only. It is not a legal compliance score, a conformity assessment, or a model performance score.

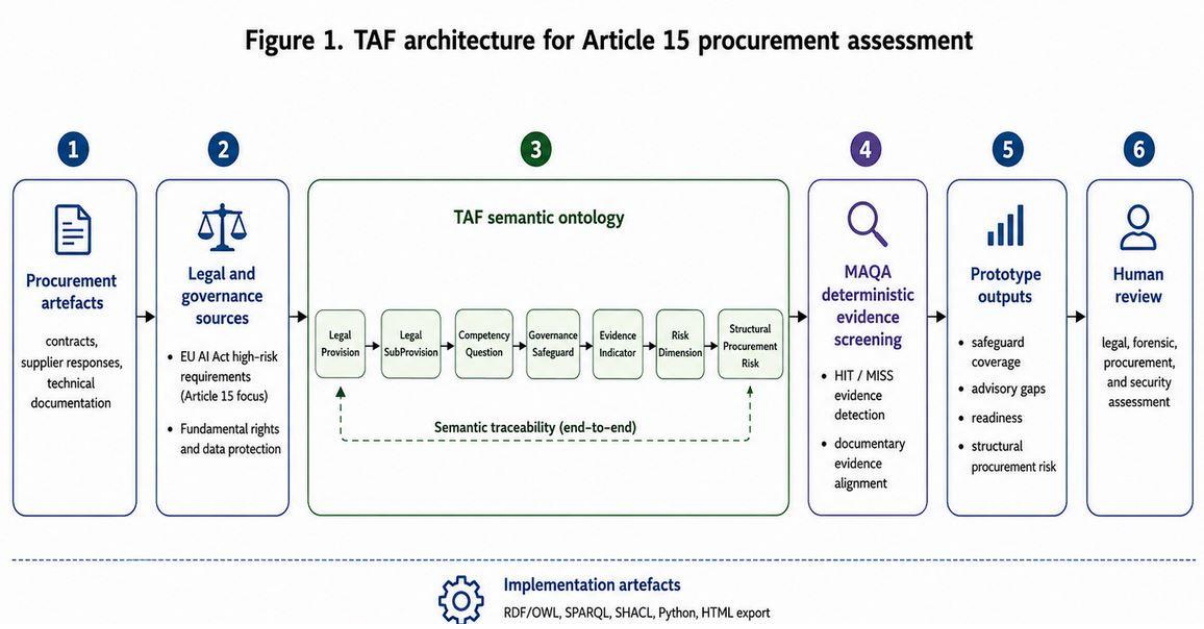


Figure 1: TAF architecture for Article 15 procurement assessment. The diagram shows the flow from procurement artefacts and legal-governance sources into the TAF semantic ontology, where semantic model concepts for legal provisions, legal sub-provisions, competency questions, governance safeguards, evidence indicators, risk dimensions, and structural procurement-risk outputs are linked through an end-to-end traceability model. The arrows show traceability and order-of-assessment links; provenance/derivation of generated artefacts is modelled separately using PROV-O. The ontology output is then screened by MAQA to produce safeguard coverage, advisory gaps, readiness, and structural procurement-risk outputs for human legal, forensic, procurement, and security review.

5. The TAF Framework

5.1. Framework Overview

TAF has three main layers: semantic governance, deterministic screening, and reporting. The semantic layer represents legal provisions, sub-provisions, competency questions, safeguards, evidence indicators, risk dimensions, and structural procurement risks. The screening layer evaluates supplier questionnaires, tender responses, contract clauses, model cards, validation reports, technical and cybersecurity documentation, data processing agreements, and AI governance statements. The reporting layer supports procurement, legal, forensic, security, and risk teams by producing traceability tables, safeguard coverage, advisory gaps, and follow-up actions such as requesting evidence, escalating legal review, imposing contract conditions, or withholding approval.

5.2. Core Semantic Model

Table 1 summarises the core semantic model.

The implemented ontology validation confirms these semantic links and reports 12 legal provisions, 33 legal sub-provisions, 31 competency questions, 12 governance safeguards, 3 evidence indicator categories, 4 risk dimensions, 7 structural procurement risk classes, 3 active contract instances, and 12 full-chain rows.

5.3. Mapping Article 15 to Procurement Safeguards

Article 15 is operationalised through procurement evidence requirements only after applicability has been considered. TAF does not decide that a system is high-risk; intended purpose and applicability rationale are treated as inputs for human legal review before Article 15-related evidence assessment

Table 1
Core TAF ontology classes.

Semantic model concept	Role in TAF
Legal provision (`LegalProvision`)	Represents a legal or governance source, such as the EU AI Act Article 15.
Legal sub-provision (`LegalSubProvision`)	Represents a granular requirement within a legal provision.
Competency question (`CompetencyQuestion`)	Defines the assessment question that scopes a safeguard.
Governance safeguard (`GovernanceSafeguard`)	Represents an operational procurement safeguard.
Evidence indicator (`EvidenceIndicator`)	Represents documentary or textual evidence of a safeguard.
Risk dimension (`RiskDimension`)	Aggregates safeguard evidence into a procurement risk dimension.
Structural procurement risk (`StructuralProcurementRisk`)	Represents the final risk-oriented procurement output.

Table 2
Article 15 mapping to procurement safeguards.

Article 15 requirement	Procurement interpretation	Example evidence indicators	Forensic risk if absent
Accuracy	Supplier defines task-specific accuracy and performance levels for the declared intended purpose.	Accuracy metric, precision, recall, error rate, validation report.	Unsupported reliance on inaccurate outputs.
Robustness	The tool performs reliably under foreseeable forensic conditions.	Robustness test, stress test, adversarial input, failure mode.	Weak performance under noisy or adversarial evidence conditions.
Cybersecurity	The tool protects models, evidence data, logs, prompts, and outputs.	Vulnerability, hardening, penetration test, access control, encryption.	Compromised evidence integrity or exposure of sensitive data.
Lifecycle consistency	Supplier maintains performance and security over time.	Monitoring, versioning, update, retraining, change control.	Degradation after deployment or an unreviewed model change.
Limitations	Supplier discloses known limitations and unsuitable uses.	Limitation, intended purpose, failure condition, instructions for use.	Over-trust in outputs beyond the validated forensic context.

is interpreted. Where Article 15 applies, documentary evidence is assessed in relation to the declared intended purpose rather than as context-independent performance claims. Related Article 13 transparency and instructions-for-use requirements and Article 14 human-oversight requirements are treated as supporting procurement context where relevant, not as separate MAQA compliance determinations.

Accuracy is mapped to declared task-specific metrics and interpreted against the system's declared intended purpose. For example, if a tool classifies documents as relevant or irrelevant, procurement documentation should specify the metric, benchmark data, validation method, false-positive and false-negative thresholds, and limitations. Robustness is assessed by testing under foreseeable forensic conditions, such as noisy data, incomplete logs, corrupted files, multilingual communications, adversarial prompts, and changing operational environments. Cybersecurity is mapped to controls protecting the AI system, model, evidence data, logs, prompts, and outputs.

5.4. Mapping Fundamental Rights to Procurement Requirements

The framework links fundamental rights to procurement safeguards because rights risks arise from technical and organisational design choices. Privacy and data protection are operationalised through safeguards for data minimisation, purpose limitation, retention, access control, logging, and secure processing [4]. Non-discrimination is operationalised through bias testing, representative validation data, human review, and disparate impact analysis [3]. Fair procedures and effective remedies are operationalised through transparency, documentation, explainability, evidence traceability, and reviewability.

These safeguards matter in both in-house and public-sector investigations because affected people may not understand how AI tools analyse or prioritise their data. Procurement documentation should therefore require auditability, human review, proportional use, and a clear route for challenging or correcting outputs. A Fundamental Rights Impact Assessment under Article 27 is not presumed to be mandatory for every procurement: where Article 27 is potentially applicable, TAF treats the presence or absence of a rights assessment as a documentary review signal subject to human legal confirmation of the relevant deployer, use case, and high-risk classification.

6. Prototype Evaluation

6.1. Evaluation Material

The prototype evaluation uses a static exported prototype report generated by the TAF/MAQA checker. The paper does not rely on a password-protected or mutable frontend as a cited source. Instead, the reported figures are treated as outputs from a frozen evaluation export.

The exported report contains three relevant components. The coverage interpretation component defines HIT/MISS and coverage. The use-case demonstration component provides the three procurement use-case results. The validation component provides the active-run and ontology-validation summaries. This separation prevents the paper from mixing method explanations, contract-level outputs, validation checks, and legal interpretation into a single results section.

The online prototype demonstration provides the static exported checker output used to derive the reported use-case summaries, while the WIDOCO documentation provides an inspectable HTML view of the ontology terms and structure. The prototype was implemented as a Python 3 command-line pipeline using `download.py` for public procurement dataset retrieval and filtering and `mcc_ai_checker.py` for TAF/MAQA screening. The implementation uses JSON/JSONL and CSV/XLSX intermediates, RDFLib for RDF/OWL serialisation to RDF/XML, SHACL shapes, SPARQL query files, WIDOCO/WebVOWL ontology documentation, and static HTML reporting. Source code, ontology files, shapes, queries, and release artefacts are made available through the persistent project URI in Table 5.

Figure 2 summarises the evaluation workflow used to move from the frozen prototype export to the workshop-paper results. Figure 3 then gives a fuller end-to-end view of the implemented TAF artefact pipeline, including dataset preparation, semantic assessment, checker outputs, ontology export, and provenance.

6.2. Procurement Use Cases

The prototype retains three public-procurement cases because they represent distinct AI value-chain configurations: provider-to-deployer procurement with GPAI involvement, provider-to-deployer procurement without GPAI, and direct GPAI procurement by a deployer. Cases 1 and 2 are police-related procurements; Case 3 tests direct GPAI procurement in a different public-sector setting. They are documentary demonstrations rather than matched in-house case studies. Their inclusion does not establish that the procured systems are AI-enabled in operation, legally high-risk, compliant, or suitable for a particular forensic use. Instead, the cases test whether TAF can expose evidence gaps and applicability questions across different procurement structures.

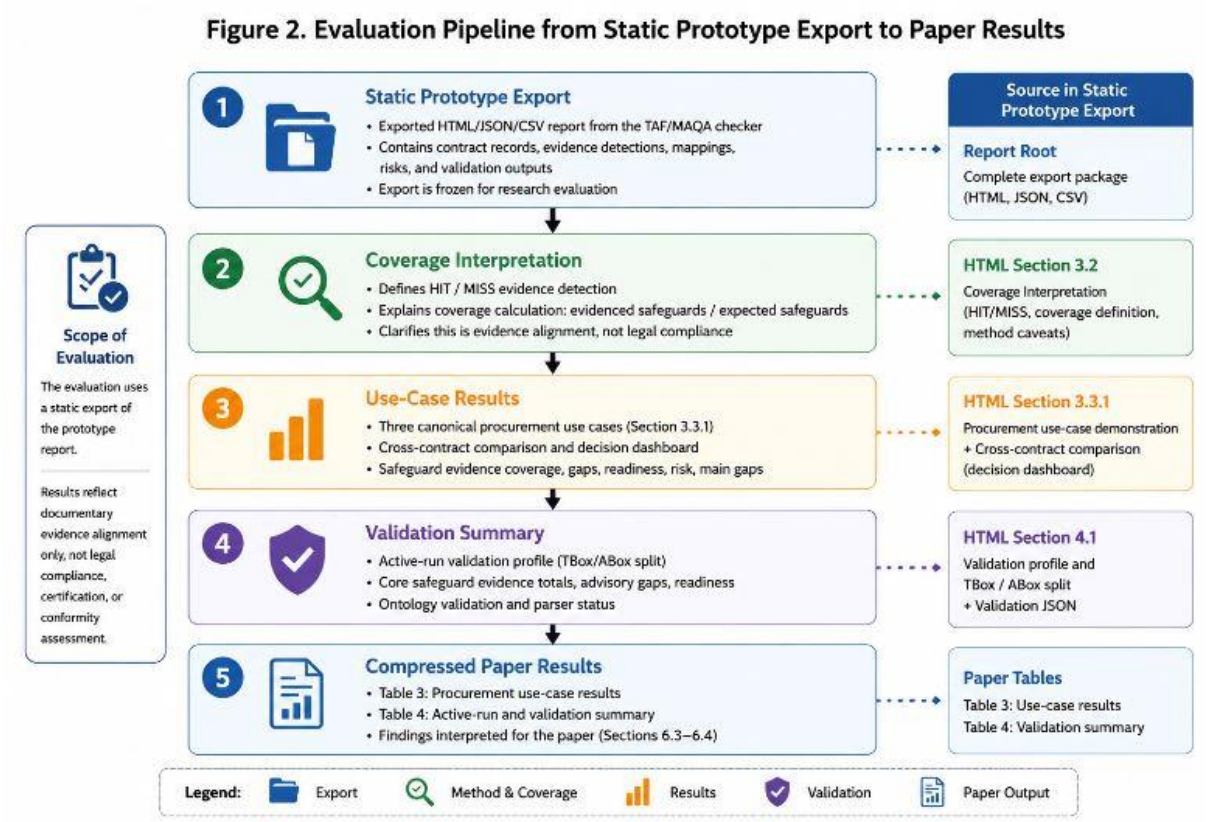


Figure 2: Evaluation pipeline process used to move from the frozen prototype export to use-case summaries, ontology validation checks, and workshop-paper results. The pipeline separates coverage interpretation, use-case extraction, active-run validation, ontology-validation outputs, and results synthesis so that contract-level evidence, semantic validation, and legal interpretation remain distinct.

Table 3
Prototype procurement use-case results.

Use case	Scenario and contract	Evidence and readiness	Structural risk and main review gap
Use Case 1	AI provider to AI deployer using GPAI. Contract 1: 709560452, Provision of CRIMES HQ Cellebrite Licence and Support.	Core safeguards: 8/8 (100.0%). Advisory gaps: 2. Readiness: 100/100.	Moderate structural risk. Gap: fundamental-rights assessment / Article 27 applicability review.
Use Case 2	AI provider to AI deployer without GPAI. Contract 2: 708269450, Service Police Cybercrime Centre.	Core safeguards: 7/8 (87.5%). Advisory gaps: 2. Readiness: 88/100.	High structural risk. Gap: forensic chain of custody and data integrity.
Use Case 3	AI deployer procures GPAI directly. Contract 3: York Global Initiative for Safe Autonomy Building.	Core safeguards: 8/8 (100.0%). Advisory gaps: 1. Readiness: 100/100.	Moderate structural risk. Gap: Annex III applicability and sector-specific obligations.

6.3. Use-Case Results

The results distinguish full documentary coverage from a record with a missing core safeguard. Use Case 2 shows that high coverage does not necessarily imply low structural procurement risk: 7 of 8 safeguards were evidenced, but the missing chain-of-custody and data-integrity safeguard is central to forensic reliability. The resulting label is a governance output from the analysed documentation, not a finding that the system is a legally defined high-risk AI system.

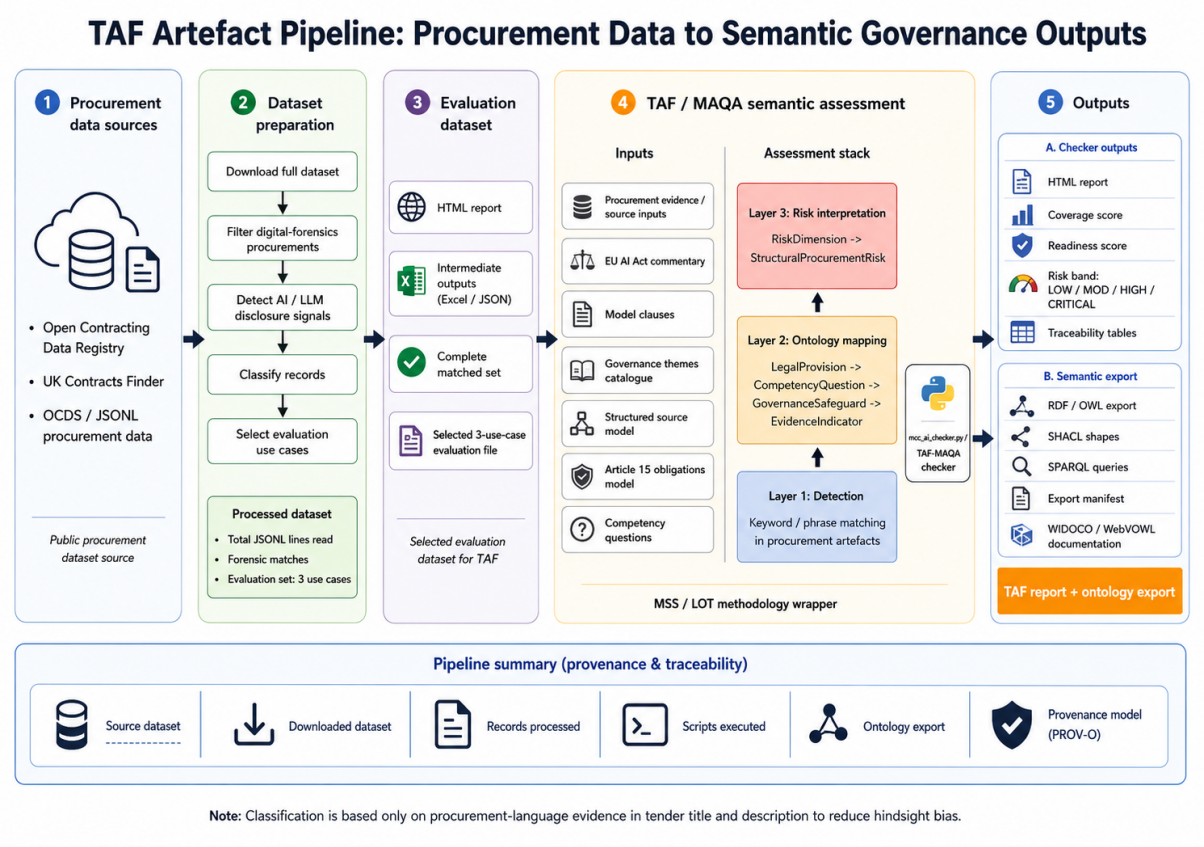


Figure 3: End-to-end TAF artefact pipeline from procurement data to semantic governance outputs. The figure shows how public procurement data from the Open Contracting Data Registry and UK Contracts Finder are downloaded and filtered for digital forensic procurements, screened for AI/LLM disclosure signals, organised into an evaluation dataset, and processed by the TAF/MAQA checker. The checker combines procurement evidence, legal and governance sources, ontology inputs, and Article 15 competency mappings to produce coverage and readiness scores, risk bands, traceability tables, and semantic exports including RDF/OWL, SHACL, SPARQL, WIDOCO, and provenance artefacts.

6.4. Active-Run and Artefact Validation Summary

Table 4 is generated from `mtalo_complete_protege_ready_validation.json`.

Table 4
Active-run and semantic validation summary.

Metric	Result
Active procurement contract groups	3
Core safeguards evidenced	23/24
Core safeguard evidence coverage	95.8%
Advisory review gaps	5
Average readiness	96.0/100
High structural procurement-risk outputs	1
Moderate structural procurement-risk outputs	2
Low structural procurement-risk outputs	0
Ontology validation status	PASS
RDF/XML parser status	PASS
Full-chain rows	12

The active-run result should be read as evidence coverage, not empirical classification accuracy. The framework has not yet been evaluated against independently reviewed `expected_status` labels. Therefore, precision, recall, F1, false positives, false negatives, and classification accuracy remain the evaluation metrics for future work.

6.5. Structural Procurement Risk Interpretation

StructuralProcurementRisk is the framework's final risk-oriented output, derived from evidence coverage, safeguard strength, advisory gaps, and aggregated RiskDimension results. The labels low, moderate, and high describe procurement-documentation risk only. They are not the AI Act's legal category of a "high-risk AI system" and should not be treated as a legal conclusion.

The full traceability tables are useful for audit and debugging, but they are too large for a workshop paper. The paper therefore presents `StructuralProcurementRisk` as a compact output column and refers to Figure 1 and Table 1 for the semantic traceability model rather than repeating the same class sequence inline. This is sufficient to show the semantic contribution without overwhelming the reader.

7. Discussion

TAF contributes to trustworthy AI procurement by operationalising legal and governance obligations into procurement-level safeguards. Article 15 requires appropriate levels of accuracy, robustness, and cybersecurity, but procurement teams need practical ways to ask for and review supplier evidence [1]. TAF provides a structured vocabulary and traceability model for doing so.

The framework is standards-oriented in three respects. First, it translates legal obligations into reusable procurement controls. Second, it creates machine-readable assessment artefacts using RDF/XML, SHACL, SPARQL, JSON, Python outputs, and HTML reporting [9, 11, 12]. Third, it supports standardised evidence requests, allowing buyers to ask suppliers for accuracy metrics, robustness testing, cybersecurity controls, logging capabilities, data governance documentation, human oversight controls, and evidence of the impact on fundamental rights, using a consistent vocabulary. This complements IEEE 3119-2025 by providing a semantic representation of evidence requests and review gaps for the specific case of rights-sensitive AI-based digital forensic procurement [8].

For in-house digital forensic teams, the framework supports earlier identification of tool risks. It helps teams ask whether an AI tool is suitable for a specific forensic purpose, rather than relying on generic AI performance claims. Missing audit logs affect the chain of custody and the reviewability

of records. Missing accuracy metrics affect reliability. Missing oversight controls increase the risk of automation bias. Missing data governance controls increase privacy and data protection risk.

The framework supports collaboration between legal, procurement, forensic, security, and risk teams: legal reviewers examine applicability and source provisions; forensic reviewers assess integrity and traceability; procurement teams request supplier evidence; security teams assess controls; and risk owners decide whether to seek further evidence, add contractual conditions, escalate, or withhold approval. TAF provides a basis for treating legal mappings, procurement evidence, and risk outputs as governed information assets, including source provenance, versioning, review status, ownership, and change history. Because the mappings, SHACL constraints, SPARQL queries, taxonomies, and screening rules are declarative, they can be revised when legislation, standards, guidance, contracts, or system versions change without rewriting the whole pipeline.

8. Limitations and Future Work

The framework has six main limitations. First, it evaluates procurement documentation rather than deployed-system behaviour, so a claim may not reflect technical implementation. Second, it does not perform model-level statistical validation. Third, it does not determine legal compliance or high-risk status; these depend on legal interpretation, intended purpose, deployment facts, contractual terms, and guidance. Fourth, deterministic lexical screening may miss equivalent wording or detect weak evidence. Fifth, the evaluation uses selected public procurement artefacts whose descriptions do not prove actual AI functionality. Sixth, the legal mappings and reports have not yet been independently validated through a legal-expert review or practitioner user study.

Future work should formalise applicability concepts in the ontology, extend the legal and forensic sources, expand the evidence taxonomy, strengthen SHACL and SPARQL validation, and test larger datasets. Maintainability should be demonstrated through controlled updates reflecting Regulation (EU) 2026/1744, subsequent Commission guidance, CEN/CENELEC JTC 21, ISO/IEC JTC 1/SC 42, ENISA guidance, and sector-specific rules [2]. Independent legal-expert review and a practitioner study should assess the correctness, clarity, and decision usefulness of the mappings and reports. Independently reviewed `expected_status` labels are also required to measure precision, recall, F1, and false-positive behaviour.

9. Online Resources and Artefact Availability

To support review, transparency, and reuse of the semantic artefacts, the project provides online resources for the prototype outputs, ontology documentation, and public repository landing page. These resources make the framework inspectable beyond the static paper and support open-access expectations in semantic web venues.

Table 5
Online resources and artefact availability.

Resource	Availability
Prototype demo	http://trinners.com/TAF_NXDG2026_Artefact_v1/mcc_ai_checker_output.html. Static exported TAF/MAQA checker output showing coverage, advisory gaps, readiness, and StructuralProcurementRisk outputs.
Ontology documentation	http://trinners.com/TAF_NXDG2026_Artefact_v1/Widoco/. WIDOCO-generated documentation for the TAF ontology, including human-readable ontology documentation and term-level inspection.
Project repository / persistent URI	https://w3id.org/taf. Persistent landing page for the open-source repository and release artefacts, including source code, ontology files, SHACL shapes, SPARQL queries, documentation, and versioned exports.

Where access to live or sensitive prototype material cannot be fully open during review, the paper links to static documentation and exported demonstration outputs so reviewers can inspect the semantic model, mappings, and reporting structure without relying on a mutable or private system.

10. Conclusion

This paper proposes TAF, a semantic governance framework for assessing Article 15-related safeguards in AI-based digital forensic procurement. It separates the human legal question of applicability from the technical task of screening procurement documentation, and links legal provisions to competency questions, safeguards, evidence indicators, risk dimensions, and structural procurement-risk outputs.

The framework is paired with MAQA, a deterministic screening method that checks procurement artefacts for explicit evidence of safeguards. The implemented artefact includes RDF/XML, SHACL, SPARQL, JSON taxonomy files, a Python screening pipeline, and HTML reporting. Validation confirms the presence of the semantic traceability links and demonstrates that procurement artefacts can be assessed for evidence coverage in a reproducible way.

TAF does not certify compliance, determine high-risk status, replace legal judgement, or benchmark model performance. Its contribution is to make applicability questions and AI-forensic procurement safeguards explicit, traceable, queryable, and reviewable. Rights-sensitive procurement should not rely on generic claims of accuracy, security, or compliance; it should require structured evidence and multidisciplinary human review.

11. Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT (OpenAI) to assist with the creation and refinement of figures and visual materials. The authors reviewed and validated all AI-assisted outputs, contributed all substantive content, and take full responsibility for the publication's content.

References

[1] European Parliament and Council, Regulation (EU) 2024/1689 Laying Down Harmonised Rules on Artificial Intelligence, Official Journal of the European Union, 2024.

[2] European Parliament and Council, Regulation (EU) 2026/1744 Amending Regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as Regards the Simplification of the Implementation of Harmonised Rules on Artificial Intelligence, Official Journal of the European Union, L 2026/1744, 2026. URL: https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng.

[3] European Union, Charter of Fundamental Rights of the European Union, Official Journal of the European Union, 2012.

[4] European Parliament and Council, Regulation (EU) 2016/679: General Data Protection Regulation, Official Journal of the European Union, 2016.

[5] E. Casey, Digital Evidence and Computer Crime: Forensic Science, Computers and the Internet, 3 ed., Academic Press, London, 2011.

[6] ISO/IEC, ISO/IEC 27037:2012: Information technology – Security techniques – Guidelines for Identification, Collection, Acquisition and Preservation of Digital Evidence, International Organization for Standardization, 2012.

[7] Public Buyers Community, Procurement of AI, Updated EU AI Model Contractual Clauses, Public Buyers Community, 2025.

[8] IEEE Standards Association, IEEE 3119-2025: IEEE Standard for the Procurement of Artificial Intelligence and Automated Decision Systems, IEEE Standards Association, 2025. URL: https://standards.ieee.org/ieee/3119/10729/.

[9] R. Cyganiak, D. Wood, M. Lanthaler, RDF 1.1 Concepts and Abstract Syntax, W3C Recommendation, 2014.

[10] P. Hitzler, M. Krötzsch, B. Parsia, P. F. Patel-Schneider, S. Rudolph, OWL 2 Web Ontology Language Primer, W3C Recommendation, 2012.
[11] S. Harris, A. Seaborne, SPARQL 1.1 Query Language, W3C Recommendation, 2013.
[12] H. Knublauch, D. Kontokostas, Shapes Constraint Language (SHACL), W3C Recommendation, 2017.
[13] European Parliament and Council, Directive 2014/24/EU on Public Procurement, Official Journal of the European Union, 2014.
[14] M. Poveda-Villalón, A. Fernández-Izquierdo, M. Fernández-López, R. García-Castro, LOT: An Industrial Oriented Ontology Engineering Framework, Engineering Applications of Artificial Intelligence 111 (2022) 104755.
[15] T. Lebo, S. Sahoo, D. McGuinness, et al., PROV-O: The PROV Ontology, W3C Recommendation, 2013. URL: https://www.w3.org/TR/prov-o/.