2.5.3 AI Opacity

Return to 2.5 Why a Reference Architecture Is Needed

Artificial intelligence-assisted techniques are increasingly used in reporting systems to extract, classify, enrich, and analyze reported information. These techniques can improve scalability, processing speed, and operational efficiency, particularly when systems process large volumes of unstructured or semi-structured information.

Artificial intelligence-assisted processing also introduces forms of opacity, nondeterminism, variability, and uncertainty that differ from traditional rule-based processing.

In many current implementations, systems incorporate artificial intelligence outputs directly into downstream interpretation or validation pipelines without a clear architectural distinction among:

- What a reporting party asserted
- What a system observed or extracted
- What an artificial intelligence component inferred
- What a governed interpretation process concluded
- What a validation or conformance process assessed

Model behavior, confidence thresholds, training data, model versions, prompt logic, retrieval context, and configuration may remain implicit or tool-specific. As a result, organizations may find it difficult to explain how a particular output was produced, determine which assumptions influenced it, or reassess the output under revised semantic definitions or governance rules.

When systems introduce artificial intelligence-assisted processing without architectural constraints, they may amplify ambiguity rather than reduce it. Common failure modes include:

- Derived values appearing authoritative without adequate provenance
- Probabilistic outputs being treated as asserted facts
- Confidence measures being omitted or interpreted inconsistently
- Model or configuration changes altering behavior without visible governance
- Inference and interpretation are being collapsed into a single opaque result
- Historical outputs are becoming difficult to reproduce or reassess
- Tool-specific behavior becoming embedded as de facto semantic authority

These conditions undermine traceability, auditability, comparability, reproducibility, and evidentiary continuity, particularly in regulated reporting environments where organizations must explain outcomes and reconstruct historical interpretations.

AI opacity is therefore not an unavoidable property of artificial intelligence techniques. It is an architectural failure mode that arises when systems integrate artificial intelligence capabilities without explicit governance of their semantic roles, authority, inputs, outputs, versions, uncertainty, and permissible uses.

A Reference Architecture provides the structure needed to incorporate artificial intelligence-assisted components while preserving clear distinctions among reported facts, observed facts, inferred results, interpreted facts, and validation outcomes.

This structure requires explicit treatment of:

- Source and input provenance
- Model and configuration identity
- Version context
- Confidence and uncertainty measures
- Prompt and retrieval context, where applicable
- Authority assigned to the output
- Relationships to downstream interpretation and validation
- Audit trails and historical reconstruction

The Federated Data Interpretation Systems Reference Architecture (FDIS-RA) constrains how systems use artificial intelligence outputs rather than prescribing which artificial intelligence techniques, models, products, or platforms they employ.

This approach supports responsible adoption of artificial intelligence-assisted capabilities without surrendering semantic control, interpretive accountability, implementation flexibility, or governance authority.
