Operationalizing Article 6 of the EU AI Act
Detecting Latent High-Risk Classification via Ontological Reasoning

Problem Context

Article 6 of the EU AI Act classifies systems as High Risk based not only on how they are deployed, but on what they are capable of doing under Annex III conditions.

Core Risk

In manufacturing contexts, perimeter security drones may bear biometric identification capability at the hardware level, even when that capability is disabled in software. This creates latent regulatory risk.

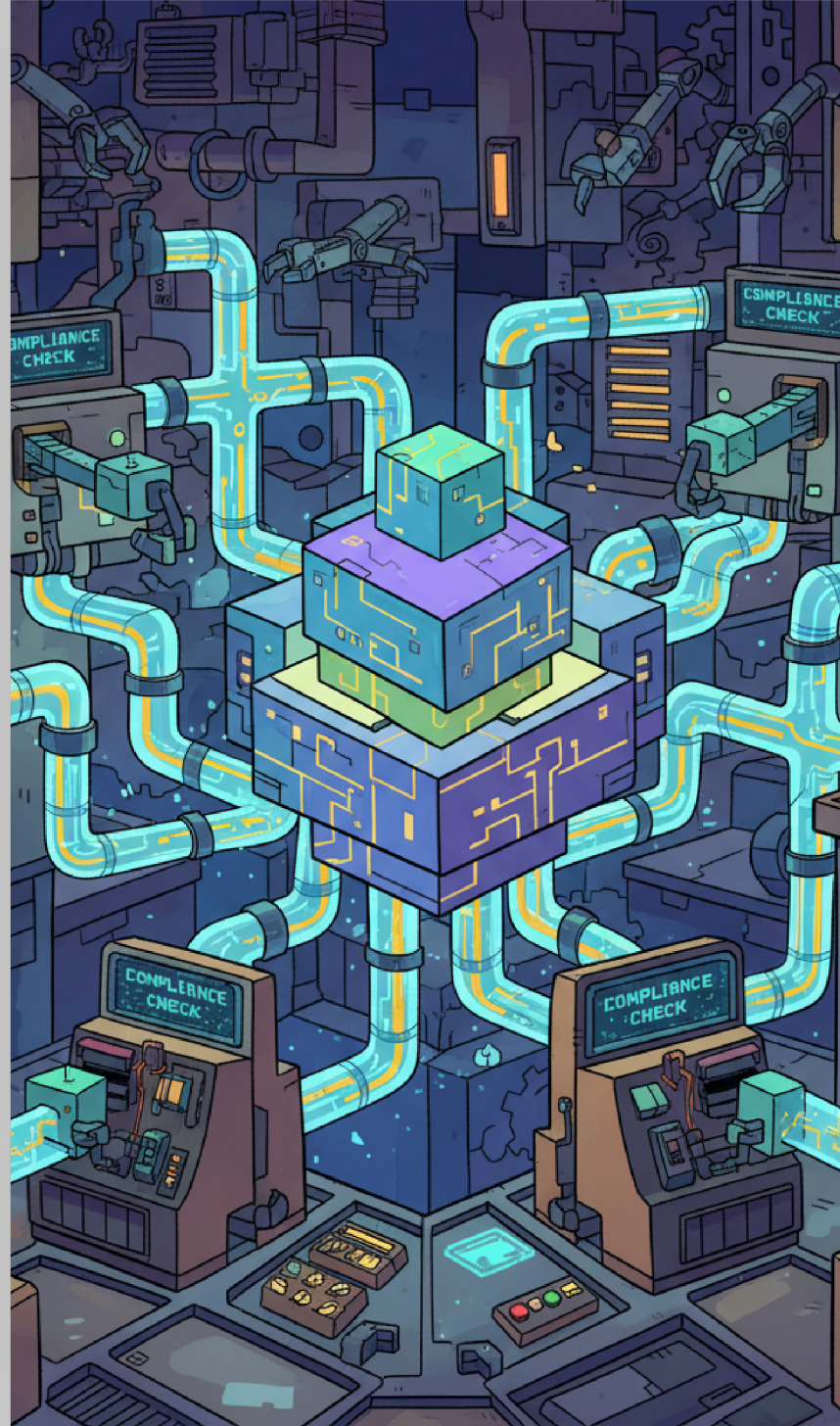
Technical Challenge

Legal criteria are expressed in natural language universals (“biometric”, “remote”, “public space”), while system specifications exist as technical artifacts. Manual review and static documentation cannot reliably track this mismatch.

Our Approach

ARCO operationalizes Article 6 by representing legal conditions as ontological constraints and reasoning over system dispositions, enabling deterministic high-risk classification based on what a system is, not merely what it is configured to do.

PROJECT 6 FOCUS: DATA & DESIGN PATTERN EXPANSION THROUGH LEGAL-DOMAIN REASONING, BFO GROUNDING, AND OWL-BASED INFERENCE



Why We Created a Synthetic AI System (Sentinel-ID)

The Challenge

Real AI systems have proprietary hardware and closed documentation, making them impractical to fully audit or understand in an academic or open setting.

The Data Problem

Real compliance data is not publicly available, preventing transparent testing of regulatory logic.

Our Focus

This project tests **reasoning correctness**, not vendor claims. We need to prove the compliance logic works.

Why Sentinel-ID?

It contains explicit latent biometric capability and its components are fully observable and modeled.

The Advantage

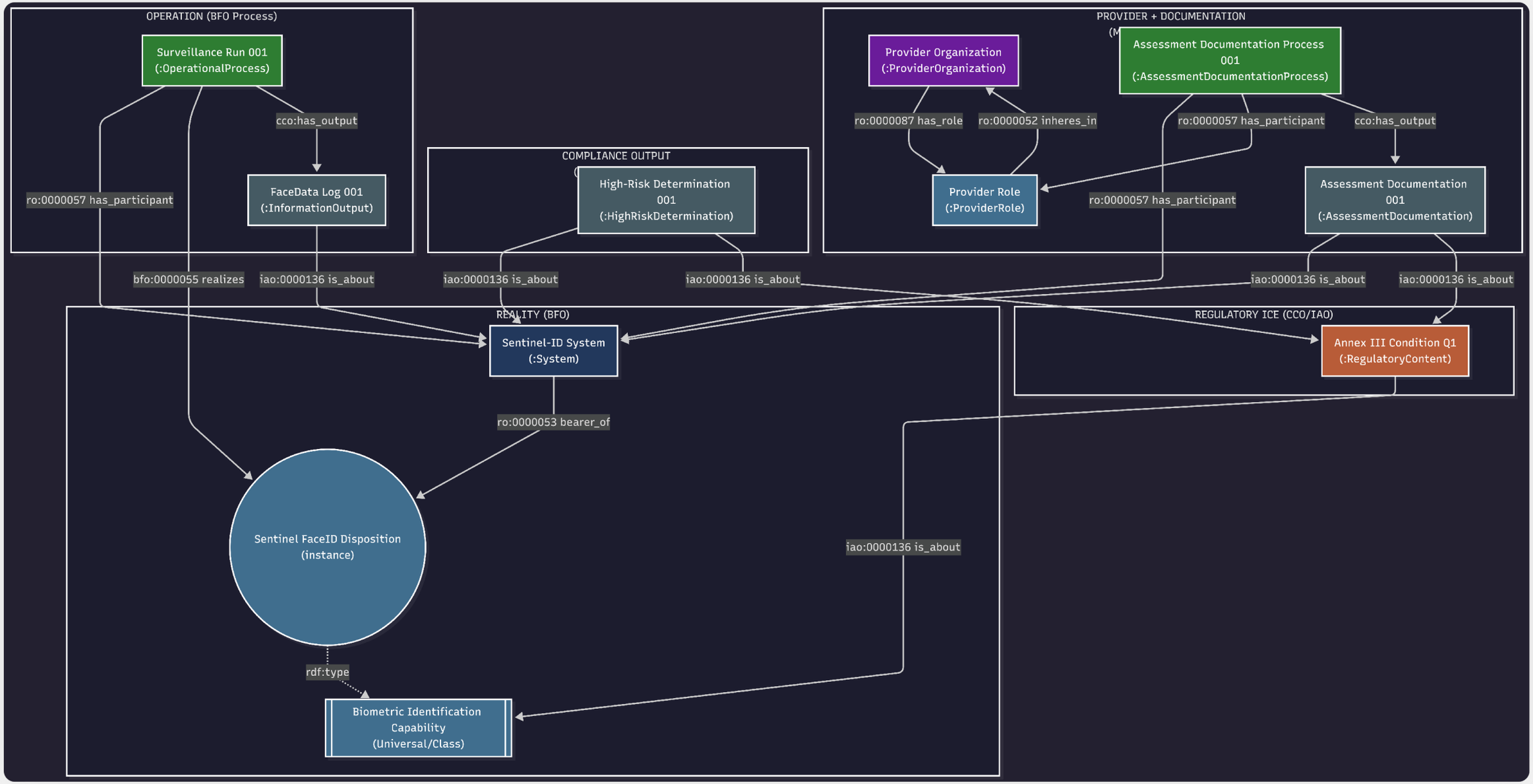
Sentinel-ID allows us to evaluate Annex III classification criteria deterministically with complete transparency.

Important Clarification

Sentinel-ID is **not a guess** about a real system. It is a controlled reference system designed to validate the compliance reasoning pipeline.

A synthetic system gives us the control and observability needed to rigorously test compliance logic.

ARCO: Ontological Compliance Model (System Overview)



Why the EU AI Act Must Be Modeled, Not Coded

The EU AI Act is written in natural language for human interpretation, not as executable software instructions. To enable computational reasoning about compliance, we must first represent the law correctly.

Understanding the nature of legal text:

- The Act uses natural language to communicate regulatory intent to people
- Article 6 describes *how systems should be evaluated*, not computational procedures
- Annex III lists c*onditions and categories*, not executable logic rules
- Direct translation of legal language into code risks fundamental misinterpretation

The key principle:

In this project, the law serves as a *reference model* that constrains and guides decisions—not as logic that autonomously makes decisions.

This approach respects what law actually is: a framework for human judgment, not an algorithm.

Why There Are Two EU AI Act Models

The EU AI Act serves two distinct roles, requiring two distinct models.

Model 1: Classification Structure

- Represents Article 6 and Annex III classification content
- Captures lists, conditions, and classification criteria
- Answers: "What makes a system High Risk?"

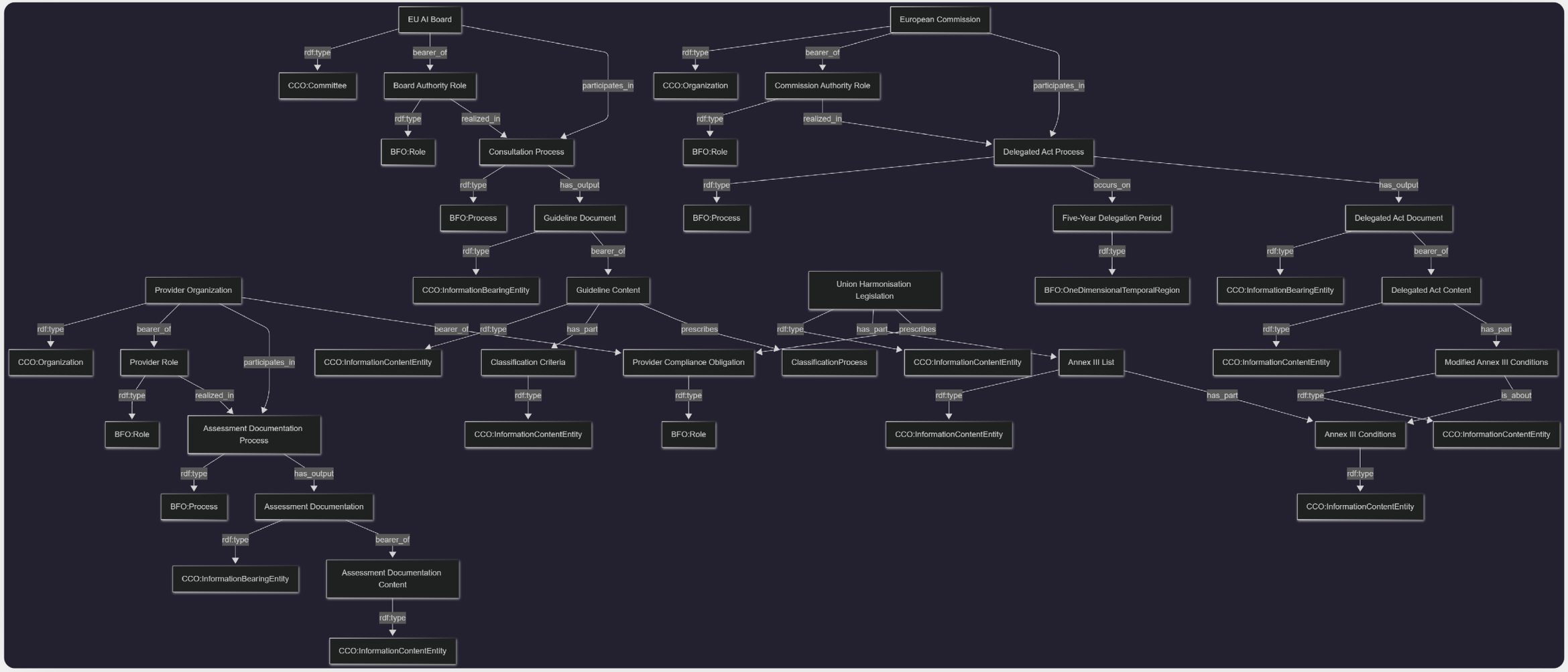
Model 2: Governance & Authority

- Represents who defines, updates, and enforces the rules
- Models EU AI Board, Commission, and delegated acts
- Answers: "Who has authority over these rules, and how they change over time?"

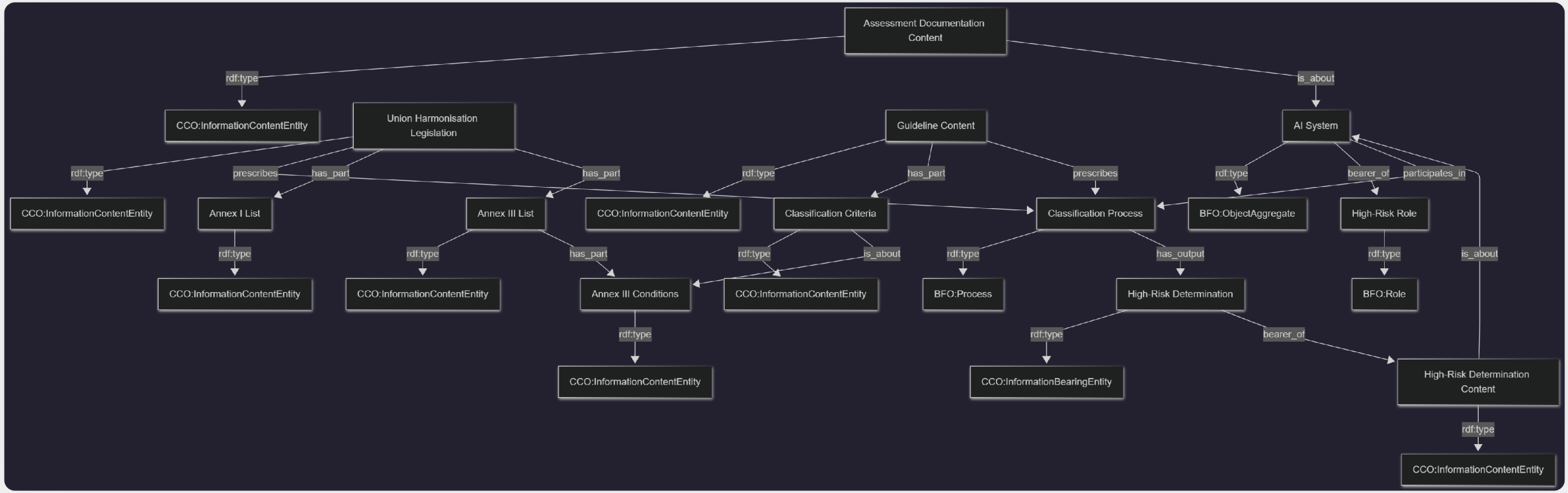
Why they are separate:

- Classification logic should not change when governance changes
- Governance can evolve without breaking system reasoning

Key takeaway: Separating regulatory structure from governance preserves clarity, realism, and long-term extensibility.



Article 6 EU AI Data Act


Assessment Documentation Content
rdf:type
CCO:InformationContentEntity
Union Harmonisation Legislation
rdf:type
CCO:InformationContentEntity
prescribes
has_part
Annex I List
rdf:type
CCO:InformationContentEntity
has_part
Annex III List
rdf:type
CCO:InformationContentEntity
has_part
Annex III Conditions
rdf:type
CCO:InformationContentEntity
Guideline Content
rdf:type
CCO:InformationContentEntity
has_part
Classification Criteria
rdf:type
CCO:InformationContentEntity
is_about
prescribes
Classification Process
rdf:type
BFO:Process
has_output
High-Risk Determination
rdf:type
CCO:InformationBearingEntity
bearer_of
is_about
AI System
rdf:type
BFO:ObjectAggregate
bearer_of
High-Risk Role
rdf:type
BFO:Role
participates_in
is_about
High-Risk Determination Content
rdf:type
CCO:InformationContentEntity

EU AI Act: Formal Regulatory Source Model

Article 6 & Annex III represented as ontological classification criteria, not executable logic

Regulatory Source

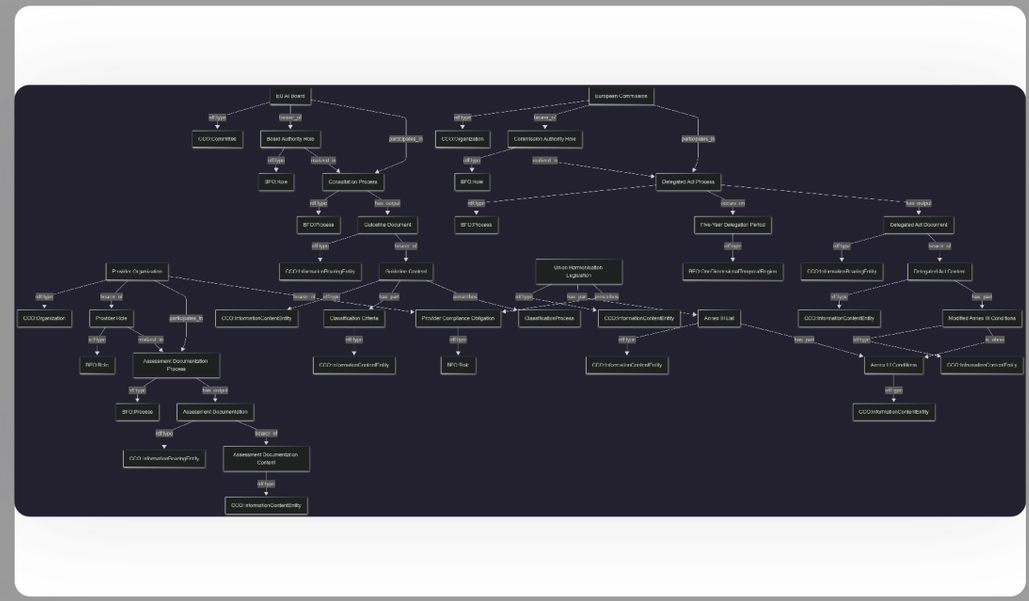
EU AI Act provisions modeled as IAO / CCO Information Content Entities (Annex III Lists, Conditions, Classification Criteria)

Classification Mechanism

Article 6 specifies how systems are evaluated, not outcomes → Referenced by a Classification Process, not procedural rules

Constraint Role

Regulation governs classification. High-risk determinations occur only when asserted system facts satisfy Annex III conditions



This regulatory model is referenced by the ARCO ontology to enable deterministic high-risk classification without embedding legal text into code.

ARCO: Ontological Compliance Model (System Overview)

Realist Core

BFO/CCO grounding for representing reality level entities. The Sentinel-ID System possesses a Biometric Identification Capability modeled as a Disposition that inheres in the system. When this disposition is realized in operational processes (e.g., Surveillance Run 001), those processes produce concrete outputs such as FaceData Log 001, establishing a formal link between latent capabilities and their manifestation in real-world operations.

Regulatory Layer

IAO usage for modeling legal texts as Information Content Entities, not procedural rules. Regulatory provisions such as Annex III Condition Q1 are represented as informational entities that are the subject of compliance determinations. The High-Risk Determination 001 is about both the regulatory content and the system's operational outputs, enabling derived conclusions rather than predictive classifications.

Governance Extension

v3 governance graph for provider traceability and documentation. The Provider Organization bears a Provider Role that participates in the Assessment Documentation Process 001. This process produces Assessment Documentation 001 that is formally linked to the system, its determinations, and the regulatory provisions, ensuring full traceability of compliance artifacts back to responsible organizational entities.

The Interpretability Gap

Law describes risk conditions; systems describe what they are built to do.

Regulatory Reality

Expressed as ***natural language*** universals, not **schemas**

Defines risk by ***capability***, not **current behavior**

Applies to ***classes of systems***, not **individual deployments**

VS

Technical Reality

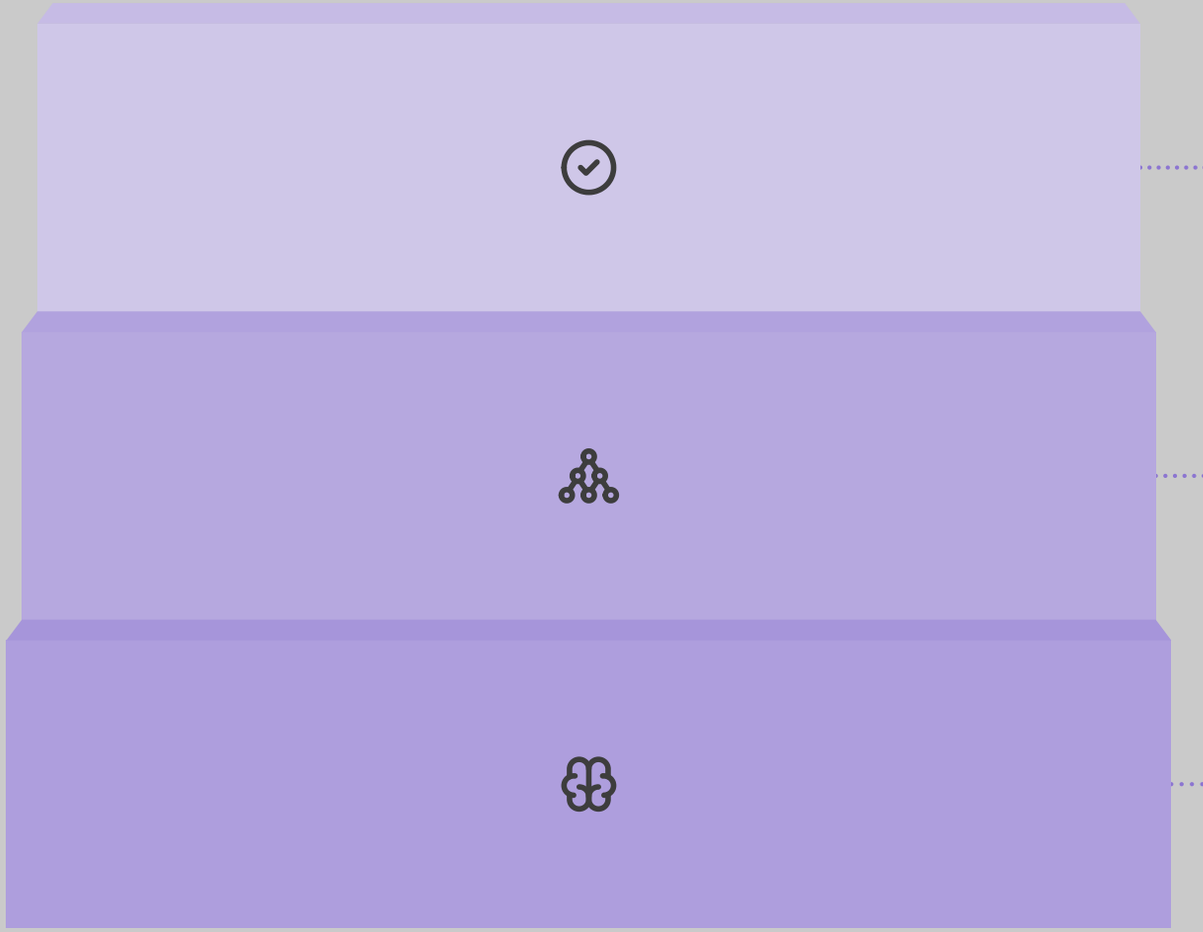
Systems are **physical assemblies** with latent capabilities

Hardware may support restricted functions **even when disabled**

Risk is **not explicitly labeled** anywhere in system data

Traditional ETL pipelines operate at the schema level, not the semantic level. Purely generative AI is dangerous—LLMs might hallucinate compliance. **Latent capability creates latent liability.** We need a system that converts probabilistic signal (text) into deterministic fact (logic).

Architecture & Project Alignment



Stage 3: Reasoning (Logic)
HermiT infers OWL restrictions to infer High-Risk

Stage 2: Representation (Symbolic)
BFO grounding represents extracted claims as axioms

Stage 1: Interpretation (Neuro)
LLM extracts capabilities from unstructured text

3-Stage Architecture

Uses **LLM for recall**, **Ontology for precision**, **Logic for proof**

LLMs as Candidate Generators, Not Judges

RAW OUTPUT

Probabilistic

LLM proposes candidate capabilities based on legal and technical text

Noisy

Output may include irrelevant or invalid relationships

Ungrounded

LLM output is a hypothesis, not a compliance conclusion

No Direct Touch

LLM never touches compliance conclusions directly—only proposes possibilities

GROUNDED

Typed

Candidates are mapped to explicit BFO classes and relations

Validated

Only assertions consistent with BFO constraints are retained

Constrained

If LLM hallucinates a relationship that doesn't fit 'bearer_of' constraint, system rejects it

Asserted

Resulting Turtle is machine-readable and reasoner-ready

LLMs interpret text. Ontologies decide truth.

Grounding Risk in Reality

Risk

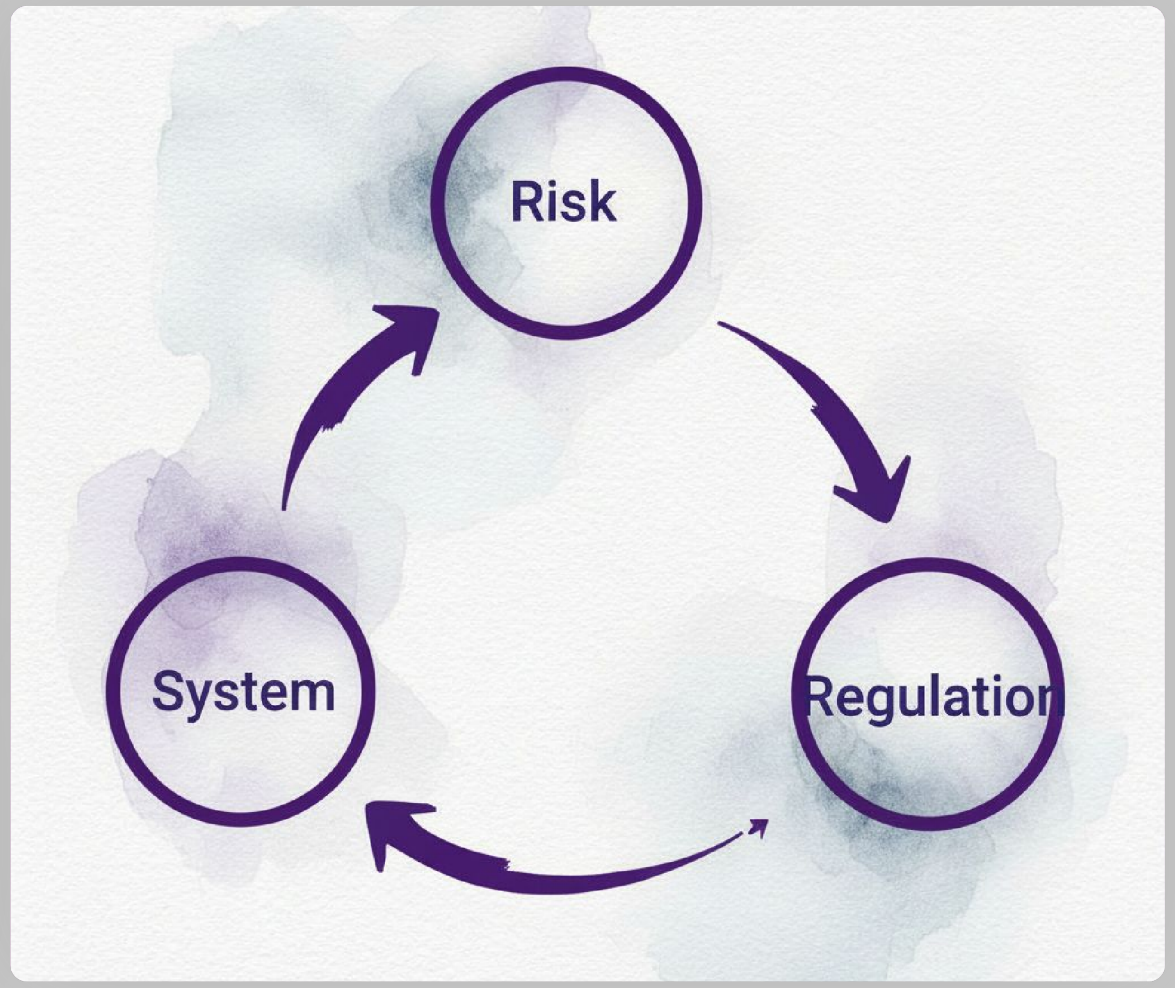
BFO – Inherent capability of a system to cause harm, existing even when latent or inactive

System

BFO - Physical bearer of the disposition; hardware carries capability independent of software state

Regulation

IAO – Legal information content that refers to and constrains system dispositions



Risk modeled as **Disposition** detects potential illegality based on what a machine **IS**, not just what it is **DOING** – biometric capability exists in hardware even when software is disabled.

Deterministic Inference

The Logic Layer: No AI involved—formal logic only

OWL Class Definition: HighRiskDetermination ≡ System AND (bearer_of SOME BiometricDisposition)

Reasoner Execution: HermiT computes class membership via OWL entailment

Logical Necessity - If premises are satisfied, the conclusion must hold

Auditability - This is proof, not approximation

Operational Validation & Governance

SHACL

Structure Validation

Ensures the knowledge graph satisfies required structural constraints before reasoning.
Invalid or incomplete compliance graphs are rejected upstream, preventing false classifications.

SPARQL

Deterministic Traceability

Enables exact, machine-checkable queries over the compliance graph, supporting audit, explanation, and liability tracing without statistical uncertainty.

- The governance extension links inferred risk determinations to real-world responsibility
- Each High-Risk classification is traceable from the system, through the provider role, to the responsible organization
- This produces a justification structure suitable for audit, regulatory review, and liability analysis

Machine-checkable validation ensures compliance integrity; governance network provides full accountability chain.

Why Glass-Box Compliance Scales

Healthcare

Clinical AI systems embed latent diagnostic and biometric capabilities. Regulatory compliance depends on provable alignment between system capabilities and approved clinical use, not post-hoc model explanations

Finance

Financial systems contain dormant decision pathways that can activate regulatory obligations. Glass-box compliance separates what a system can do from what it is authorized to do, enabling deterministic audit and accountability

Defense

Defense systems are regulated based on capability, not deployment intent. Compliance requires traceable proof that classified capabilities are detected and governed before operational use

The neuro-symbolic pattern scales because it enforces a hard boundary between probabilistic interpretation and deterministic validation, enabling compliance reasoning wherever latent capability creates regulatory exposure

From Black-Box to Glass-Box

Thank you.