CRSA-1 EU Edition

Compositional Runtime Safety
Attestation Protocol

EU AI Act Compliance Profile
for Multi-Model AI Systems

Auburn Patent Family Fields
Ryan Fields
2026

Application Layer Document
Auburn Governance Stack
Document 1 of the CRSA-1 EU Compliance Series

FIRST IN EU COMPLIANCE SERIES

Intellectual Property (IP) Declaration

The methods, logic structures, compositional risk taxonomies, conformity assessment methodologies, and Article 25(4) contract templates contained in this work are the sole property of **Ryan Fields**.

Public License (Non-Commercial)

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

- **Academic Use:** Researchers may share and use this framework for non-commercial academic purposes, provided full attribution is given to Ryan Fields.
- **No Derivatives:** No modifications or adaptations of the compositional risk taxonomies, obligation mappings, or conformity assessment methodologies are permitted without express written consent.

Commercial Prohibition

Commercial use of this framework is strictly prohibited. This includes, but is not limited to:

- Use within proprietary AI governance, risk, or compliance (GRC) platforms.
- Integration into commercial conformity assessment tooling or notified body evaluation methodologies.
- Use by consultancies, law firms, or systems integrators for billable client engagements without a commercial license.
- Incorporation into enterprise quality management systems (QMS) for EU AI Act compliance purposes.

UncleBroFields@proton.me

fieldsryanchristopher@gmail.com

EU Compliance Series

This document is Document 1 of the CRSA-1 EU Compliance Series. Subsequent editions will provide:

- Sector-specific compliance overlays (DORA/Financial Services, MDR/Medical Devices, NIS2/Critical Infrastructure).
- Member State regulatory transposition profiles.
- Detailed implementation parameters, threshold values, and conformity assessment test vectors.

Commercial licensing terms apply uniformly across the series. All rights reserved for forthcoming editions.

Honest Framing

This specification provides compositional safety methodology aligned to the obligations of Regulation (EU) 2024/1689 (the EU AI Act). It does not guarantee regulatory approval, behavioral safety, or immunity from enforcement action. It provides the technical infrastructure—specifically, the compositional risk taxonomy, obligation mapping, conformity assessment methodology, and contract templates—that no harmonized European standard currently offers for composed AI systems.

This document operates within the honest framing principle of the Auburn Governance Stack: compositional safety attestation provides probabilistic risk reduction and accountability infrastructure, not behavioral guarantees. This is analogous to financial auditing, which certifies process compliance without guaranteeing future solvency.

Where this document identifies obligations as "implicit" or "derived" from the AI Act text, this reflects the author's interpretive analysis of the regulation. Definitive interpretation rests with the European Commission, the AI Office, and ultimately the Court of Justice of the European Union.

Contents

1 Scope and Regulatory Context

1.1 Purpose

Regulation (EU) 2024/1689—the Artificial Intelligence Act—establishes the world's first comprehensive, risk-based regulatory framework for AI systems placed on the European market. Its obligations span risk management, technical documentation, logging, transparency, human oversight, accuracy, robustness, and cybersecurity. These obligations apply to all high-risk AI systems regardless of internal architecture.

The regulation was drafted, however, for a paradigm of single, bounded AI systems with identifiable intended purposes, discrete training datasets, and deterministic operational envelopes. The production landscape has moved beyond this paradigm. Multi-model pipelines, Retrieval-Augmented Generation (RAG) architectures, ensemble voting systems, cascade routing configurations, and multi-agent workflows built on protocols such as the Model Context Protocol (MCP) now constitute the dominant deployment pattern for enterprise AI in regulated sectors.

These composed systems present a category of risk that no existing harmonized European standard addresses: **compositional safety**—the phenomenon whereby individual AI components may satisfy compliance requirements in isolation yet generate emergent, unpredictable risks when composed into interacting architectures. Error propagation across pipeline stages, bias amplification through ensemble aggregation, semantic drift across model boundaries, and state corruption in multi-agent workflows represent failure modes that cannot be detected, measured, or mitigated using single-model evaluation methodologies.

This document provides the compositional safety framework that the AI Act's obligations require but that the current standardization programme does not deliver. It maps every relevant obligation in Regulation 2024/1689 to specific compositional safety requirements, establishes a risk taxonomy for composed AI systems, provides architecture-specific compliance profiles for the five dominant multi-model patterns, defines a conformity assessment methodology for composed systems, and publishes the first Article 25(4) contract template for multi-vendor AI pipelines.

1.2 Scope

This specification applies to AI systems that satisfy **both** of the following conditions:

1. The system is classified as high-risk under Article 6 of Regulation (EU) 2024/1689, whether through Annex I product integration (Article 6(1)) or Annex III use-case designation (Article 6(2)).

2. The system is **composed**—that is, it incorporates two or more AI models, general-purpose AI models, or AI-driven components that interact during inference, where the output of one component influences the behaviour of another. This includes, but is not limited to:

   - Retrieval-Augmented Generation (RAG) pipelines combining retrieval, reranking, and generation models.
   - Agentic systems employing planner, executor, critic, or tool-calling model configurations.
   - Ensemble architectures aggregating outputs from multiple models through voting, averaging, or weighted consensus mechanisms.
   - Cascade and routing systems directing inputs to different models based on classifier or threshold logic.
   - Multi-agent systems connecting models to external tools, data sources, or other AI systems through protocols such as MCP.

1.3 Exclusions

This document does **not** address:

- Single-model AI systems operating independently without compositional interaction. These systems are adequately served by the harmonized standards under development within CEN-CENELEC JTC 21.
- General-purpose AI model provider obligations under Chapter V (Articles 51–56) of the AI Act, except insofar as GPAI models are integrated as components within composed high-risk systems.
- Prohibited AI practices under Article 5 of the AI Act.
- AI systems classified as minimal risk or limited risk (transparency-only obligations under Article 50).
- Obligations arising exclusively from the General-Purpose AI Code of Practice, unless those obligations intersect with high-risk system composition requirements.

1.4 The Regulatory Vacuum

As of March 2026, the regulatory infrastructure intended to operationalize the AI Act's high-risk requirements exhibits a structural gap regarding compositional safety. This gap is not speculative; it is documented across the standardization programme, the enforcement pipeline, and the regulatory guidance landscape.

1.4.1 No Harmonized Standard Addresses Compositional Safety

CEN-CENELEC Joint Technical Committee 21 (JTC 21), operating under Standardisation Request M/593 (May 2023, amended by M/613), is responsible for drafting the harmonized standards that grant a "presumption of conformity" under Article 40. All standards under development assume a monolithic, single-system architecture:

- **prEN 18286** (AI Quality Management System)—the most advanced standard, having entered Enquiry on 30 October 2025—provides no guidance on quality management across multi-model compositions, RAG pipelines, or agentic architectures.
- **prEN 18228** (AI Risk Management), carrying over 400 reconsideration requests, contains no methodology for assessing emergent risks from component interaction, cascading failures in pipelines, or risk propagation across model boundaries.
- **prEN 18229-1** (Logging, Transparency, Human Oversight) provides no framework for distributed logging across pipeline stages, decision attribution to specific components, or human oversight of autonomous multi-agent workflows.
- **prEN 18229-2** (Accuracy and Robustness) provides no methods for measuring end-to-end accuracy of composite systems and no framework for robustness degradation across pipeline stages.
- **prEN 18282** (Cybersecurity), undergoing comprehensive redraft following European Commission review, has not reached Enquiry. OWASP AI Exchange contributors have noted that it does not address agentic scenarios.
- **prEN 18284** (Dataset Quality and Governance) provides no guidance on data governance for dynamic RAG retrieval corpora or data quality management across multiple model inputs.

- **prEN 18285** (Conformity Assessment Framework) provides no methodology for conformity assessment of systems composed of multiple AI components from different providers.

The ten ISO/IEC standards mapped by JTC 21—including ISO/IEC 42001, 23894, 24028, 24029, 25059, 5338, and 22989—similarly provide no coverage of compositional safety. ISO/IEC 24029, which addresses neural network robustness, operates exclusively at the single-network level; the insight that compositional robustness cannot be derived from the product of individual robustness measures is absent from the standard.

1.4.2 No Notified Bodies Possess Multi-Model Assessment Capability

Under Articles 28–39, Member States are responsible for designating Notified Bodies to conduct third-party conformity assessments required by Annex VII. As of March 2026, no AI Act-specific Notified Bodies have been formally designated. The likely candidates—TÜV SÜD, TÜV Rheinland, Bureau Veritas, DNV, SGS, and DEKRA—are developing AI assessment capabilities, but none has published a methodology for evaluating multi-model AI systems. The standards against which these bodies would benchmark their assessments are themselves incomplete.

1.4.3 The Digital Omnibus Contains No Composition Language

The Digital Omnibus on AI (COM(2025) 836 final), published 19 November 2025, proposes a "stop-the-clock" mechanism decoupling enforcement dates from the fixed August 2026 timeline, with long-stop dates of 2 December 2027 for Annex III systems and 2 August 2028 for Annex I systems. However, the proposal contains no specific provisions for multi-model, multi-agent, or composed AI systems. As of March 2026, the Omnibus is navigating the ordinary legislative procedure with trilogue negotiations expected mid-2026. Adoption before 2 August 2026 is uncertain.

1.4.4 The AI Office Has Not Published Article 25(4) Contract Terms

Article 25(4) mandates that providers of high-risk AI systems and third parties supplying components used or integrated in those systems specify, by written agreement, the necessary information, capabilities, technical access, and other assistance required for compliance. The AI Office is empowered to develop voluntary model terms for such contracts but has not published any. The eleven guidelines announced by the Commission on 4 December 2025—including guidance on value chain responsibilities and substantial modification—have not been released.

1.4.5 Institutional Recognition of the Gap

The compositional safety gap is not identified solely by this specification. The Future Society's June 2025 report, *Ahead of the Curve: Governing AI Agents under the EU AI Act*, concluded that "technical standards under development will likely fail to fully address risks from agents." The Bruegel think tank noted that "agentic AI systems. . . are not covered" by the Digital Omnibus and that "the Act is already running behind the AI technology curve." The European Commission's own FAQ acknowledges that "developments related to AI agents are recent and fast evolving" and that "the European Commission's regulatory considerations are only preliminary at this stage."

No EU-specific compositional AI safety specification exists. The closest approximations—the Agent Behavioral Contracts (ABC) academic framework, the Cloud Security Alliance MAESTRO framework, and OWASP's Agentic Top 10—are not aligned to EU regulatory obligations, do not provide conformity assessment mapping, and do not address the full scope of AI Act requirements.

1.5 Enforcement Timeline

The obligations addressed by this specification operate under the following enforcement schedule:

Obligation Category	Trigger	Date
Annex III high-risk requirements (Chapter III, Section 2)	Fixed date under current AI Act text	**2 August 2026**
Annex III high-risk requirements (if Omnibus adopted)	6 months after Commission readiness decision, or long-stop	**2 December 2027** (long-stop)
Annex I high-risk requirements	Fixed date under current AI Act text	**2 August 2027**
Annex I high-risk requirements (if Omnibus adopted)	12 months after Commission readiness decision, or long-stop	**2 August 2028** (long-stop)
Product Liability Directive transposition	Member State transposition deadline	**9 December 2026**
Cyber Resilience Act full applicability	Fixed date	**11 December 2027**

The penalty structure for non-compliance with high-risk requirements reaches **€15 million or 3% of total worldwide annual turnover**, whichever is higher (Article 99(4)). Non-compliance with the AI Act additionally triggers a rebuttable presumption of defectiveness under Article 10(2)(b) of the revised Product Liability Directive (2024/2853).

1.6 Relationship to the Auburn Governance Stack

This document is an application-layer profile within the Auburn Governance Stack, consuming attestation evidence through the MAI-1 composition waist as defined in CRSA-1 (Auburn Clause AI-9). MAI-1 defines the canonical interface through which lower-layer governance guarantees—platform attestation, model state invariants, and provenance binding—are delivered as verifiable evidence. This EU Edition maps that evidence infrastructure to the specific obligations of Regulation (EU) 2024/1689.

The base CRSA-1 specification provides the architecture-neutral compositional safety protocol. This EU Edition provides the regulation-specific compliance profile. The two documents are designed for joint operation but may be referenced independently.

1.7 EU Compliance Series

EU Compliance Series — Forthcoming

This document establishes the compositional safety framework at the EU-wide level, mapping obligations under Regulation (EU) 2024/1689 to composed AI system requirements. Subsequent editions of the CRSA-1 EU Compliance Series will provide:

- **Sector-specific compliance overlays** addressing the compounding obligations of DORA (financial services), MDR (medical devices), and NIS2 (critical infrastructure) when applied to multi-model AI systems.
- **Member State regulatory transposition profiles** reflecting national implementation of the AI Act, NIS2, and the Product Liability Directive across key jurisdictions.
- **Detailed implementation parameters**, including threshold values for compositional risk metrics, conformity assessment test vectors, and architecture-specific monitoring configurations.

Sector-specific editions are anticipated in Q3–Q4 2026.

2 Article-by-Article Obligation Mapping

This section maps every provision in Regulation (EU) 2024/1689 that explicitly or implicitly requires compositional safety analysis for multi-model AI systems. Each obligation is classified by composition relevance:

- **Explicit**—the article text directly references multi-component interaction, third-party integration, or system composition.
- **Implicit**—the obligation cannot be satisfied for a composed system without compositional analysis, even though the text does not reference composition directly.
- **Derived**—the obligation applies at the system level and, when the system is composed, necessarily requires evaluation of inter-component behaviour to demonstrate compliance.

For each obligation, the mapping identifies the specific compositional requirement, the harmonized standard gap, and the CRSA-1 mechanism that addresses the gap. Where the regulation text is paraphrased, the paraphrase is faithful to the enacted text of Regulation (EU) 2024/1689 as published in the Official Journal (OJ L, 2024/1689, 12.7.2024).

2.1 Article 3 — Foundational Definitions

The AI Act's definitional framework does not explicitly address composed systems. It does, however, establish the interpretive foundation upon which every downstream obligation operates. Four definitions are structurally significant for multi-model architectures.

Article 3(1) — Definition of "AI System"

Text (paraphrased): An AI system is a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment, and that infers from its input how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.

Composition Relevance	Derived
Compositional Implication	A composed pipeline that collectively infers outputs from inputs constitutes a single AI system under this definition. The definition does not require that inference occur within a single model; a system comprising a retriever, a reranker, and a generator that jointly produce an output is one AI system. Recital 12 reinforces this reading: "AI systems can be used on a stand-alone basis or as a component of a product."
Standards Gap	ISO/IEC 22989 (AI Concepts and Terminology), including its GenAI amendment, does not define multi-model systems, composite AI, pipelines, or agentic architectures. The vocabulary for composed AI systems does not exist within the harmonized standards programme.
CRSA-1 Mechanism	Section 3 of this specification establishes a compositional risk taxonomy that provides the definitional vocabulary for composed AI system types, failure modes, and interaction patterns.

Article 3(3) — Definition of "Provider"

Text (paraphrased): A provider is the entity that develops an AI system or a general-purpose AI model, or that has one developed, and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.

Composition Relevance	Explicit
Compositional Implication	The entity assembling a multi-model pipeline and deploying it under their name is the provider of the composed system, regardless of whether they developed the individual component models. An enterprise that integrates an open-weight foundation model from Vendor A, a vector database from Vendor B, and custom orchestration logic is the provider of the resulting system. This reading is reinforced by Article 25 (Section 2.11).
Standards Gap	No harmonized standard provides guidance on determining provider status in multi-vendor composed systems. prEN 18286 (QMS) addresses organizational responsibilities but does not define how provider obligations distribute across a multi-vendor AI value chain.
CRSA-1 Mechanism	Section 5 provides a conformity assessment methodology that includes provider determination criteria for composed systems. Appendix B provides Article 25(4) contract templates that operationalize information-sharing obligations between providers and component suppliers.

Article 3(23) — Definition of "Substantial Modification"

Text (paraphrased): A substantial modification is a change to an AI system after its placing on the market or putting into service which is not foreseen or planned in the initial conformity assessment and as a result of which the compliance of the AI system with the Chapter III, Section 2 requirements is affected, or which results in a modification to the intended purpose for which the AI system has been assessed.

Composition Relevance	Explicit
Compositional Implication	Replacing a component model within a composed pipeline meets this definition if the change was not foreseen in the initial conformity assessment and affects system-level compliance with accuracy, robustness, bias, or other Chapter III requirements. The continuous evolution inherent in multi-model architectures—component model updates, retrieval corpus changes, tool additions in MCP-based systems—creates persistent substantial modification risk. Article 43(4) provides a mitigation: changes explicitly pre-documented as "predetermined changes" in the technical documentation (Annex IV, point 2(f)) and assessed at the time of conformity assessment do not constitute substantial modifications.
Standards Gap	No harmonized standard defines the threshold at which a component change in a composed system constitutes a substantial modification. The GPAI guidelines specify that a downstream modifier becomes a new GPAI model provider if modification compute exceeds one-third of the original model's training compute, but no equivalent threshold exists for composed system component changes.
CRSA-1 Mechanism	Section 5 defines substantial modification triggers specific to each composed architecture type and provides a predetermination strategy for anticipated component changes.

Article 3(63) — Definition of "General-Purpose AI Model"

Text (paraphrased): A general-purpose AI model is an AI model that is trained with a large amount of data using self-supervision at scale, that displays significant generality, and that can be integrated into a variety of downstream systems or applications.

Composition Relevance	Explicit
Compositional Implication	The definition explicitly contemplates the composition relationship: GPAI models are designed for integration into downstream systems. When a GPAI model serves as a component within a composed high-risk system, the system provider must obtain sufficient technical documentation from the GPAI provider (per Article 53 and Annex XII) to satisfy system-level compliance obligations. The adequacy of this upstream documentation for compositional safety analysis remains unresolved.
Standards Gap	The General-Purpose AI Code of Practice addresses GPAI provider obligations but does not specify the documentation requirements necessary for downstream compositional safety assessment.
CRSA-1 Mechanism	Section 5 specifies the minimum upstream documentation requirements that composed system providers must obtain from GPAI component suppliers to satisfy conformity assessment obligations.

2.2 Article 6 — High-Risk Classification

Article 6(1)–(2) — Classification Rules

Text (paraphrased): An AI system is high-risk if it is intended to be used as a safety component of a product covered by Annex I harmonisation legislation requiring third-party conformity assessment (paragraph 1), or if it falls within the use cases enumerated in Annex III (paragraph 2).

Composition Relevance	Implicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	The AI Act does not explicitly state that a composed system inherits the highest-risk classification of any component. Classification depends on the system-level intended purpose, not component-level risk categories. However, Recital 85 states that general-purpose AI models "may be used as high-risk AI systems by themselves or be components of other high-risk AI systems." If any part of a composed system functions as a safety component of a regulated product, the entire system falls under Article 6(1). For Annex III classification, the composed system's aggregate intended purpose governs—but determining the aggregate intended purpose of a multi-agent system serving multiple functions requires compositional analysis that no guideline currently specifies. Commission guidelines on Article 6 classification were due by 2 February 2026 (Article 6(5)) but have not been published.
Standards Gap	No harmonized standard or Commission guideline provides a methodology for classifying composed AI systems where components serve different functions or where the system's intended purpose emerges from the interaction of its components.
CRSA-1 Mechanism	Section 3 provides a compositional risk taxonomy that enables system-level risk characterization. Section 4 provides architecture-specific classification guidance for each of the five dominant composed system patterns.

Article 6(3) — Derogation for Low-Risk Annex III Systems

Text (paraphrased): An Annex III AI system shall not be considered high-risk if it does not pose a significant risk of harm to health, safety, or fundamental rights, including by not materially influencing the outcome of decision-making—for example, if the system performs a narrow procedural task, improves the result of a previously completed human activity, detects decision-making patterns without replacing human assessment, or performs a preparatory task.

Composition Relevance	Derived
Compositional Implication	Invoking the Article 6(3) derogation for a composed system requires demonstrating that the *complete pipeline*—not any individual component—satisfies the derogation criteria. A component that performs a "narrow procedural task" in isolation may materially influence decision-making when its output is consumed by a downstream model that acts on it. The derogation analysis must account for how component outputs propagate through the composed system. The Digital Omnibus proposes removing the registration requirement for systems claiming this derogation, but the substantive assessment obligation remains.
Standards Gap	No harmonized standard provides a methodology for evaluating whether a composed system satisfies the derogation criteria at the system level.
CRSA-1 Mechanism	Section 3 provides the compositional risk taxonomy necessary to evaluate whether emergent pipeline behaviour exceeds the derogation thresholds.

2.3 Article 8 — Compliance Framework

Article 8(1)–(2) — System-Level Compliance

Text (paraphrased): High-risk AI systems shall comply with the requirements of Chapter III, Section 2, taking into account their intended purpose and the generally acknowledged state of the art in AI and AI-related technologies (paragraph 1). Where a product contains an AI system, the provider of that product shall be responsible for ensuring that the product is fully compliant with all applicable requirements (paragraph 2).

Composition Relevance	Derived
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Article 8 establishes a cumulative compliance obligation assessed at the system level. For composed systems, this means compliance cannot be demonstrated by aggregating individual component compliance certificates. The overall system's intended purpose governs compliance, not individual component purposes. A composed system where each component individually meets Chapter III requirements may nonetheless fail system-level compliance if the composition introduces risks not present in any component. The "state of the art" reference in Article 8(1) is significant: as compositional safety methodologies emerge, failure to apply them may itself constitute non-compliance.
Standards Gap	No harmonized standard provides a methodology for demonstrating system-level compliance of composed AI systems. prEN 18286 (QMS) addresses organizational compliance processes but not the technical evidence required to prove that a composed system satisfies Chapter III requirements as an integrated whole.
CRSA-1 Mechanism	This specification provides the system-level compliance framework for composed systems. Section 4 provides architecture-specific compliance profiles. Section 5 provides the conformity assessment methodology. Together, these deliver the technical evidence infrastructure that Article 8 demands at the system level.

2.4 Article 9 — Risk Management System

Article 9(1)–(2) — Continuous Risk Management

Text (paraphrased): A risk management system shall be established, implemented, documented, and maintained as a continuous iterative process planned and run throughout the entire lifecycle of a high-risk AI system, requiring regular systematic updating. The risk management system shall identify and analyse the known and reasonably foreseeable risks that the high-risk AI system can pose to health, safety, or fundamental rights.

Composition Relevance	Implicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	The "continuous iterative process" requirement means that every component update, model swap, retrieval corpus modification, or tool addition within a composed system triggers risk management reassessment. The "known and reasonably foreseeable risks" obligation implicitly demands compositional risk analysis: the state space of possible actions in a multi-agent system grows combinatorially with each added component, tool, or model. Risks that are foreseeable at the system level—error propagation, cascading hallucination, semantic drift—may not be foreseeable from any single component's risk profile. A model that safely refuses harmful requests in isolation may enable harmful outcomes when embedded in an agentic scaffold that chains together multiple tool calls.
Standards Gap	prEN 18228 (AI Risk Management), mapped to ISO/IEC 23894, contains no methodology for evaluating combinatorial risk pathways in composed systems. It approaches risk from a single-system governance perspective, lacking mathematical or algorithmic frameworks for assessing dynamic model-to-model interactions.
CRSA-1 Mechanism	Section 3 provides the compositional risk taxonomy—the classification of risk categories specific to composed systems—that Article 9 requires but no standard delivers. Each risk category includes affected articles, detection methodology, and mitigation patterns.

Article 9(4) — Interaction Effects

Text (paraphrased): The risk management measures referred to in paragraph 2 shall give due consideration to the effects and possible interaction resulting from the combined application of the requirements set out in Chapter III, Section 2.

Composition Relevance	Explicit
Compositional Implication	This is the closest the AI Act comes to explicitly requiring compositional risk analysis. The "effects and possible interaction" language demands that risk management evaluate not merely whether each Chapter III requirement is met individually, but whether satisfying one requirement creates tension with or undermines another. For composed systems, this obligation extends naturally: optimizing accuracy in one pipeline stage may degrade robustness in another; logging granularity sufficient for traceability may compromise data minimisation under GDPR. The interaction effects between requirements are compounded by the interaction effects between components.
Standards Gap	No harmonized standard provides a framework for evaluating interaction effects between Chapter III requirements in composed systems. prEN 18228 treats requirements independently.
CRSA-1 Mechanism	Section 3 identifies cross-requirement interaction patterns specific to composed architectures. Section 4 maps these interactions to each of the five architecture profiles, identifying where requirement tensions are most acute.

Article 9(5)–(8) — Testing

Text (paraphrased): Providers shall ensure that high-risk AI systems are tested for the purpose of identifying the most appropriate and targeted risk management measures (paragraph 5). Testing shall be suitable to achieve the intended purpose and carried out against prior defined metrics and probabilistic thresholds (paragraph 5). Providers shall eliminate or reduce risks as far as possible through adequate design and development (paragraph 6). Residual risks must be communicated to deployers (paragraph 7). Foreseeable misuse shall be considered (paragraph 8).

Composition Relevance	Derived
Compositional Implication	Testing a composed system against "prior defined metrics and probabilistic thresholds" requires compositional metrics that capture system-level behaviour, not merely component-level performance. Residual risk communication to deployers must include compositional residual risks—risks that remain specifically because the system is composed. Foreseeable misuse of a composed system includes adversarial exploitation of inter-component communication channels, retrieval corpus poisoning in RAG pipelines, and prompt injection attacks that propagate across agent boundaries.
Standards Gap	prEN 18229-2 (Accuracy and Robustness) does not provide compositional testing metrics. No standard defines probabilistic thresholds for composed system behaviour.
CRSA-1 Mechanism	Section 4 provides architecture-specific testing requirements including compositional accuracy metrics, cross-component robustness tests, and adversarial testing profiles for each architecture type.

2.5 Article 10 — Data and Data Governance

Article 10(1)–(2) — Dataset Quality Requirements

Text (paraphrased): High-risk AI systems that utilise techniques involving the training of AI models with data shall be developed on the basis of training, validation, and testing datasets that meet the quality criteria referred to in paragraphs 2 to 5 (paragraph 1). Training, validation, and testing datasets shall be subject to data governance and management practices appropriate for the intended purpose of the AI system (paragraph 2).

Composition Relevance	Implicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Data governance cascades across every component in a composed system. Each model's training, validation, and testing data must individually comply with Article 10 requirements. Additionally, the composed system's overall data pipeline must be governed: if Component A is trained on Dataset X and Component B on Dataset Y, the provider must ensure both comply individually *and* that their interaction does not create aggregate bias or errors not present in either dataset alone.
Standards Gap	prEN 18284 (Dataset Quality and Governance) provides no guidance on data governance for multi-model data pipelines where different components consume different datasets with potentially different quality characteristics, distributions, and bias profiles.
CRSA-1 Mechanism	Section 3 defines "Aggregation Bias" as a compositional risk category arising when individually compliant datasets interact through composed models to produce aggregate bias. Section 4 provides architecture-specific data governance requirements for each pipeline type.

Article 10(2)(f) — Bias Examination

Text (paraphrased): Data governance shall address examination in view of possible biases that are likely to affect the health and safety of persons, have a negative impact on fundamental rights, or lead to discrimination prohibited under Union law, especially where data outputs influence inputs for future operations (feedback loops).

Composition Relevance	Derived
Compositional Implication	In composed systems, feedback loops extend across component boundaries. A RAG pipeline where the generator's outputs are subsequently indexed by the retriever creates a cross-component feedback loop that can amplify bias iteratively. Ensemble systems where individual model biases interact through voting aggregation create non-linear bias dynamics that cannot be assessed by examining any single model's training data. The reference to "data outputs influencing inputs for future operations" is directly applicable to agentic systems where one agent's output becomes another agent's input within the same inference cycle.
Standards Gap	No harmonized standard addresses cross-component bias feedback loops or aggregation bias dynamics in ensemble or pipeline architectures.
CRSA-1 Mechanism	Section 3 defines "Aggregation Bias" and "Semantic Drift" as compositional risk categories that capture cross-component bias dynamics. Section 4 provides bias assessment methodologies specific to each architecture type.

Article 10(5) — Operational Input Data

Text (paraphrased): To the extent that it is strictly necessary for the purpose of ensuring bias detection and correction, the providers of high-risk AI systems may exceptionally process special categories of personal data, subject to appropriate safeguards.

Composition Relevance	Derived
Compositional Implication	For RAG pipelines, the retrieval corpus constitutes operational input data that is functionally distinct from model training data but equally determinative of system output quality. Article 10's governance requirements apply to this corpus, but no standard specifies what data governance for a dynamic, continuously updated retrieval corpus requires. The Digital Omnibus proposes expanding the derogation for processing special category data for bias correction across more actors—this expansion becomes particularly significant in composed systems where bias detection may require access to special category data at multiple pipeline stages.
Standards Gap	prEN 18284 does not address data governance for dynamic RAG retrieval corpora or distinguish between training data governance and operational input data governance.
CRSA-1 Mechanism	Section 4 (RAG Pipeline profile) defines data governance requirements for retrieval corpora as distinct from model training data governance, including freshness, relevance, and bias monitoring requirements.

2.6 Article 11 and Annex IV — Technical Documentation

Article 11(1) — Documentation Obligation

Text (paraphrased): The technical documentation of a high-risk AI system shall be drawn up before that system is placed on the market or put into service and shall be kept up to date. The technical documentation shall be drawn up in such a way as to demonstrate that the high-risk AI system complies with the requirements set out in Chapter III, Section 2, and to provide national competent authorities and notified bodies with the necessary information to assess compliance.

Composition Relevance	Explicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Technical documentation must demonstrate compliance of the composed system as an integrated whole. Merely appending individual component documentation does not satisfy Article 11. The documentation must show how the composition itself meets every Chapter III requirement—including how inter-component interactions affect risk management (Article 9), accuracy (Article 15), and robustness (Article 15). The "kept up to date" obligation means that every component change affecting system-level compliance triggers a documentation update, regardless of whether the change constitutes a substantial modification under Article 3(23).
Standards Gap	prEN 18286 (QMS) addresses documentation management processes but provides no templates or structural requirements for documenting composed AI systems. No harmonized standard specifies what compositional documentation must contain.
CRSA-1 Mechanism	Section 4 provides architecture-specific documentation requirements for each of the five composed system types, structured to satisfy the Annex IV elements identified below.

Annex IV

Text (paraphrased): Technical documentation shall contain a description of the elements of the AI system and of the process for its development, including recourse to pre-trained systems or tools provided by third parties and how those have been used, integrated, or modified by the provider.

Composition Relevance	Explicit
Compositional Implication	This provision directly mandates documentation of third-party AI component integration. For a composed system incorporating a GPAI model from Provider A, a retrieval engine from Provider B, and custom orchestration logic, Annex IV 2(a) requires documenting each third-party component, how it was integrated, and what modifications (if any) were applied. The provision implicitly requires the provider to have obtained sufficient technical information from upstream suppliers to produce this documentation— connecting directly to the Article 25(4) contract obligation.
Standards Gap	No harmonized standard specifies the level of detail required when documenting third-party AI component integration, the minimum upstream technical information necessary, or how to document integration when the upstream provider restricts access to model internals for trade secret protection.
CRSA-1 Mechanism	Section 5 specifies minimum upstream documentation requirements. Appendix B provides Article 25(4) contract templates that include documentation disclosure provisions.

Annex IV

Text (paraphrased): Technical documentation shall contain a description of the system architecture explaining how software components build on or feed into each other and integrate into the overall processing.

Composition Relevance	Explicit
Compositional Implication	This is the most direct compositional documentation obligation in the AI Act. The language "how software components build on or feed into each other" is an unambiguous mandate to document pipeline architecture, inter-component data flows, API boundaries, orchestration logic, and the processing sequence through which inputs are transformed into outputs across multiple models. For agentic systems, this requires documenting agent interaction protocols, tool-calling mechanisms, and state management across asynchronous agent processes. For ensemble systems, this requires documenting voting aggregation algorithms, model weighting schemes, and consensus mechanisms. The obligation exists in the enacted text. The method for satisfying it does not.
Standards Gap	CRITICAL. No harmonized standard provides templates, structural requirements, or reference architectures for documenting how AI components "build on or feed into each other." This is the paradigmatic example of a mandate without a method.
CRSA-1 Mechanism	Section 4 provides architecture-specific documentation templates for each of the five composed system types, directly satisfying the Annex IV 2(c) requirement. Each template specifies the required architectural elements, data flow documentation, and integration point characterisation.

Annex IV

Text (paraphrased): Technical documentation shall include a general description of the AI system, including how the AI system interacts with, or can be used to interact with, hardware or software, including with other AI systems, that are not part of the AI system itself.

Composition Relevance	Explicit
Compositional Implication	This provision addresses the system boundary problem. For composed systems, the boundary between "the AI system" and "other AI systems that are not part of the AI system itself" is architecturally significant and legally consequential. In an MCP-based multi-agent system, are externally connected MCP tools "part of" the AI system or "other systems"? The answer determines the scope of conformity assessment, the extent of documentation obligations, and whether connecting a new tool constitutes a substantial modification. No regulatory guidance resolves this boundary question.
Standards Gap	No harmonized standard provides criteria for determining the system boundary of a composed AI system or for documenting interactions between the system and external AI components at the boundary.
CRSA-1 Mechanism	Section 5 provides system boundary determination criteria for composed architectures, including guidance on when external tool connections fall within or outside the assessed system boundary.

Annex IV

Text (paraphrased): Technical documentation shall contain a description of the system and its performance, including any predetermined change to the high-risk AI system and its performance that has been pre-assessed at the time of the initial conformity assessment.

Composition Relevance	Explicit
Compositional Implication	This provision is the primary mitigation mechanism for the substantial modification problem in composed systems. By pre-documenting anticipated component changes—model version upgrades, retrieval corpus updates, tool additions—and assessing their impact at the time of initial conformity assessment, the provider can execute these changes without triggering re-assessment under Article 43(4). For composed systems, this requires a systematic enumeration of foreseeable component changes and a pre-assessment of the parametric range within which each change remains compliant. This is an engineering discipline that no existing standard defines.
Standards Gap	No harmonized standard provides methodology for predetermining component changes in composed systems or for specifying the parametric boundaries within which a component change remains non-substantial.
CRSA-1 Mechanism	Section 5 defines a pre-determination framework for composed systems, specifying how to document anticipated component changes, establish acceptable parametric ranges, and structure the initial conformity assessment to accommodate foreseeable evolution.

2.7 Article 12 — Record-Keeping and Automatic Logging

Article 12(1)–(2) — Logging Requirements

Text (paraphrased): High-risk AI systems shall technically allow for the automatic recording of events (logs) over the lifetime of the system (paragraph 1). The logging capabilities shall ensure a level of traceability of the AI system's functioning throughout its lifecycle that is appropriate to the intended purpose of the system (paragraph 2).

Composition Relevance	Implicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Traceability of a composed system's functioning requires logging that captures inter-component interactions, not merely individual component outputs. In a RAG pipeline, tracing a specific output back to its causal chain requires recording which documents were retrieved, how the reranker scored them, which context was presented to the generator, and how the generator produced the final output. In a multi-agent system, traceability requires recording which agent initiated an action, which tools were called, what intermediate states were produced, and how those states influenced downstream agent behaviour. Standard API token logs are insufficient: they record invocations but not the causal relationships between them.
Standards Gap	prEN 18229-1 (Logging, Transparency, Human Oversight) is not designed for multi-component systems. It provides no guidance on distributed logging across pipeline stages, causal tracing across component boundaries, or log correlation mechanisms for asynchronous multi-agent processes.
CRSA-1 Mechanism	Section 4 provides architecture-specific logging requirements for each composed system type, including minimum logging granularity, cross-component correlation identifiers, and causal chain reconstruction capabilities.

Article 12(2) — Risk Identification Through Logs

Text (paraphrased): Logging capabilities shall, in particular, enable the identification of situations that may result in the high-risk AI system presenting a risk within the meaning of Article 79(1) or in a substantial modification, and facilitate the post-market monitoring referred to in Article 72.

Composition Relevance	Derived
Compositional Implication	Identifying situations where a composed system "presents a risk" requires detecting compositional failure modes: cascading hallucination across agent boundaries, semantic drift between retrieval and generation, routing discrimination in cascade systems, and aggregation bias emergence in ensembles. These failure modes manifest in inter-component patterns, not in individual component logs. Identifying a "substantial modification" through logging requires detecting when component behaviour has drifted beyond the predetermined change parameters documented under Annex IV 2(f)—a compositional monitoring requirement that connects logging directly to conformity assessment.
Standards Gap	No harmonized standard defines the log patterns or anomaly signatures that indicate compositional failure modes in multi-model systems.
CRSA-1 Mechanism	Section 3 defines compositional failure mode signatures. Section 4 maps these signatures to logging detection requirements for each architecture type.

2.8 Article 13 — Transparency and Provision of Information

Article 13(1) — Interpretability of Output

Text (paraphrased): High-risk AI systems shall be designed and developed in such a way as to ensure that their operation is sufficiently transparent to enable deployers to interpret the system's output and use it appropriately.

Composition Relevance	Implicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Deployers cannot interpret outputs or use a system appropriately without understanding that the output is the product of multi-model processing. A credit decision produced by a RAG pipeline that retrieves context from a regulatory database, reranks by relevance, and generates a natural language assessment requires the deployer to understand that the output reflects retrieval quality, reranking logic, and generation fidelity—not merely "the model's assessment." The Act does not explicitly require disclosing multi-model architecture to deployers, but the functional requirement of "sufficient transparency to enable appropriate use" cannot be met without compositional disclosure.
Standards Gap	prEN 18229-1 provides no framework for transparency disclosure of composed system architectures. No standard specifies what compositional information deployers require to satisfy the "interpret and use appropriately" threshold.
CRSA-1 Mechanism	Section 4 provides architecture-specific transparency requirements, including minimum compositional disclosure to deployers for each system type.

Article 13(3)(b) — Capabilities and Limitations Disclosure

Text (paraphrased): Instructions of use shall contain information on the characteristics, capabilities, and limitations of performance of the high-risk AI system, including the known or foreseeable circumstances related to the use of the high-risk AI system that may lead to risks to the health, safety, or fundamental rights (point (ii)), and the technical measures put in place to facilitate the interpretation of the outputs (point (iii)).

Composition Relevance	Derived
Compositional Implication	"Capabilities and limitations" of a composed system include emergent properties that arise from composition. A RAG pipeline's capability to retrieve and synthesise information is paired with a limitation that retrieval recall is stochastic and context quality varies. An agentic system's capability to execute multi-step workflows is paired with limitations on critic model coverage and rollback capability. These are compositional capabilities and limitations that do not appear in any individual component's documentation. "Foreseeable circumstances that may lead to risks" must include compositional failure scenarios: retrieval corpus poisoning, cascading hallucination, routing discrimination.
Standards Gap	No harmonized standard specifies how to characterise the capabilities and limitations that emerge from AI system composition.
CRSA-1 Mechanism	Section 3 provides the risk taxonomy from which compositional limitations are derived. Section 4 specifies architecture-specific capability and limitation disclosures.

2.9 Article 14 — Human Oversight

Article 14(1)–(2) — Oversight Design Requirements

Text (paraphrased): High-risk AI systems shall be designed and developed in such a way as to ensure that they can be effectively overseen by natural persons during the period in which they are in use (paragraph 1). Human oversight shall aim to prevent or minimise the risks to health, safety, or fundamental rights that may emerge when a high-risk AI system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse (paragraph 2).

Composition Relevance	Implicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	"Effective oversight by natural persons" of a composed system requires oversight of the system as a whole, not merely of individual components. For multi-agent systems operating at machine speed, the oversight challenge is qualitatively different from single-model oversight: the human overseer must comprehend the interactions between agents, the state of multi-step workflows, and the implications of intermediate outputs—in real time. The "prevent or minimise risks that may emerge" language encompasses compositional risks that emerge from system operation, not merely pre-identified risks.
Standards Gap	prEN 18229-1 provides no framework for human oversight of autonomous multi-agent workflows, real-time compositional risk monitoring, or oversight scaling for systems operating across multiple concurrent agent processes.
CRSA-1 Mechanism	Section 4 provides architecture-specific human oversight frameworks, including intervention point architecture, state visibility requirements, and oversight scaling methodologies for each composed system type.

Article 14(4)(a) — Understanding Capacities and Limitations

Text (paraphrased): The measures referred to in paragraph 1 shall enable the individuals to whom human oversight is assigned to fully understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions, and unexpected performance.

Composition Relevance	Derived
Compositional Implication	"Fully understanding" a composed system requires comprehending the pipeline architecture, the role of each component, how components interact, and where compositional failure modes may emerge. This is a substantially higher cognitive burden than understanding a single model. "Detecting anomalies" in a composed system requires monitoring inter-component behaviour patterns, not merely component-level outputs. An anomaly may manifest as normal output from each individual component but abnormal correlation between component outputs.
Standards Gap	No standard specifies the training, tooling, or informational requirements for human overseers of composed AI systems.
CRSA-1 Mechanism	Section 4 specifies oversight personnel competency requirements and anomaly detection instrumentation for each architecture type.

Article 14(4)(d)–(e) — Override and Interruption

Text (paraphrased): Individuals assigned to human oversight shall be able to decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override, or reverse the output of the high-risk AI system (point (d)), and to be able to interrupt the system through a "stop" button or a similar procedure that allows the system to come to a halt in a safe state (point (e)).

Composition Relevance	Implicit
Compositional Implication	Interrupting a composed system requires halting all active components simultaneously and managing the resulting state consistency. In an agentic healthcare triage system, if a human supervisor halts the executive agent while a tool-calling subagent is actively writing an update to an Electronic Health Record via API, the interruption must prevent data corruption, complete or cleanly roll back the in-progress write, and halt downstream agent processes that depend on the interrupted state. "Reversing the output" of a composed system requires compositional rollback capability: the ability to identify and undo the effects of outputs that have propagated across multiple components, tools, and external systems. The AI Act provides no technical parameters for safe state management in interrupted multi-agent processes.
Standards Gap	CRITICAL. No harmonized standard provides a framework for safe interruption, state rollback, or output reversal in composed AI systems. prEN 18229-1 addresses human oversight for single systems and does not consider asynchronous multi-agent state management.
CRSA-1 Mechanism	Section 4 provides architecture-specific safe interruption protocols, including state consistency requirements, rollback procedures, and halt propagation mechanisms for each composed system type.

2.10 Article 15 — Accuracy, Robustness, and Cybersecurity

Article 15(1) — Appropriate Levels Throughout Lifecycle

Text (paraphrased): High-risk AI systems shall be designed and developed in such a way as to achieve an appropriate level of accuracy, robustness, and cybersecurity, and to perform consistently in those respects throughout their lifecycle.

Composition Relevance	Derived
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Accuracy, robustness, and cybersecurity are properties of the composed system as a whole. None of these properties compose linearly. Two individually accurate models may produce an inaccurate composed output when one model's errors are amplified by the other's processing. Two individually robust models may create a fragile pipeline when adversarial perturbations at one stage propagate to exploit sensitivities at the next. Two individually secure models may create an insecure system when inter-component communication channels introduce new attack surfaces. The "consistently throughout their lifecycle" requirement means these compositional properties must be maintained as components are updated, replaced, or extended— a continuous compositional monitoring obligation.
Standards Gap	prEN 18229-2 (Accuracy and Robustness) addresses single-system evaluation exclusively. ISO/IEC 24029 (Neural Network Robustness) operates at the single-network level. No standard provides methods for specifying, measuring, or monitoring compositional accuracy, robustness, or cybersecurity.
CRSA-1 Mechanism	Section 3 defines compositional failure modes that degrade accuracy, robustness, and cybersecurity at the system level. Section 4 provides architecture-specific metrics and monitoring requirements for each property across each system type.

Article 15(2) — Declared Accuracy Metrics

Text (paraphrased): The levels of accuracy and the relevant accuracy metrics of high-risk AI systems shall be declared in the accompanying instructions of use.

Composition Relevance	Derived
Compositional Implication	Declaring accuracy metrics for a composed system requires compositional accuracy specification—a methodology for deriving system-level accuracy from component-level performance characteristics and their interaction. Retrieval precision multiplied by generation accuracy does not equal RAG pipeline accuracy. Router accuracy multiplied by downstream model accuracy does not equal cascade accuracy. Ensemble accuracy exceeds individual model accuracy in expectation but may degrade under adversarial conditions that exploit the aggregation mechanism. The requirement to "declare" these metrics assumes the existence of a methodology for computing them. No such methodology exists in the harmonized standards programme. This is a specification obligation without a specification methodology.
Standards Gap	CRITICAL. prEN 18229-2 provides no methods for specifying end-to-end accuracy of composite systems. No standard defines compositional accuracy metrics or provides mathematical frameworks for deriving system-level accuracy from component performance.
CRSA-1 Mechanism	Section 4 provides architecture-specific accuracy specification frameworks, including compositional accuracy derivation methodologies for RAG, ensemble, cascade, agentic, and MCP-based systems.

Article 15(3) — Robustness Against Errors and Faults

Text (paraphrased): High-risk AI systems shall be resilient regarding errors, faults, or inconsistencies that may occur within the system or the environment in which the system operates, in particular due to their interaction with natural persons or other systems.

Composition Relevance	Explicit
Compositional Implication	The phrase "interaction with... other systems" directly encompasses inter-component interaction within composed architectures. Resilience against errors occurring "within the system" includes resilience against errors propagating between components. A retriever that returns irrelevant context is an error within the system; the generator faithfully synthesising that irrelevant context into a confident but incorrect output is error propagation within the system. Both fall within Article 15(3). The requirement extends to "inconsistencies"—a term that captures semantic drift between model boundaries, where the meaning of an intermediate representation shifts as it passes from one component to another.
Standards Gap	No harmonized standard addresses cross-component error propagation, inter-model fault tolerance, or semantic consistency verification across pipeline stages.
CRSA-1 Mechanism	Section 3 defines "Cascade Failure" and "Semantic Drift" as compositional risk categories directly addressing Article 15(3). Section 4 provides architecture-specific robustness requirements for each failure mode.

Article 15(4)–(5) — Cybersecurity Requirements

Text (paraphrased): High-risk AI systems shall be resilient as regards attempts by unauthorised third parties to alter their use, outputs, or performance by exploiting the system vulnerabilities (paragraph 4). The technical solutions aimed at ensuring the cybersecurity of high-risk AI systems shall be appropriate to the relevant circumstances and the risks, and may include measures to prevent and control for attacks including data poisoning, model poisoning, adversarial examples or model evasion, confidentiality attacks, or model flaws (paragraph 5).

Composition Relevance	Derived
Compositional Implication	Composed systems introduce attack surfaces that do not exist in single-model deployments. An attacker may inject an adversarial payload into an external database; a retrieval model fetches this data without executing it, but passes it to an executor agent that acts on the malicious instruction—a cross-component prompt injection attack. Multi-agent MCP systems are susceptible to indirect prompt injection through any connected data source or tool. Adversarial examples crafted for one component may propagate through the pipeline to exploit downstream components that were not individually vulnerable. The enumerated attack types—data poisoning, model poisoning, adversarial examples—each acquire compositional variants when models are chained. Retrieval corpus poisoning in RAG systems is a compositional form of data poisoning. Inter-agent prompt injection is a compositional form of adversarial input.
Standards Gap	prEN 18282 (Cybersecurity), currently undergoing comprehensive redraft, does not address agentic scenarios, cross-component attack vectors, or compositional vulnerability assessment. OWASP's Agentic Top 10 identifies compositional attack patterns but is not aligned to EU regulatory obligations and provides no conformity assessment mapping.
CRSA-1 Mechanism	Section 3 defines "Context Poisoning" as a compositional risk category. Section 4 provides architecture-specific cybersecurity requirements, including cross-component attack surface analysis, inter-agent communication sandboxing requirements, and retrieval corpus integrity verification for each system type.

2.11 Article 25 — Responsibilities Along the AI Value Chain

Article 25(1) — Provider Reclassification

Text (paraphrased): Any distributor, importer, deployer, or other third party shall be considered a provider of a high-risk AI system for the purposes of this Regulation and shall be subject to the obligations of the provider laid down in Article 16, if that entity: (a) puts its name or trademark on a high-risk AI system already placed on the market or put into service; (b) makes a substantial modification to a high-risk AI system that has already been placed on the market or put into service; or (c) modifies the intended purpose of an AI system, including a general-purpose AI system, which has not been classified as high-risk, in a manner which makes it high-risk.

Composition Relevance	Explicit
Enforcement Date	2 August 2026
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Article 25(1) is the provision that transforms enterprise deployers of composed systems into providers. An enterprise that assembles a pipeline using an open-weight foundation model from Vendor A, a cloud inference endpoint from Vendor B, and a proprietary retrieval corpus, wrapping these components in custom routing or orchestration logic, is legally creating a new system. Because the custom orchestration alters the system's performance envelope—determining which queries reach which model, how context is assembled, and how outputs are aggregated—the assembly qualifies as a substantial modification under Article 3(23). The enterprise is reclassified as the provider. It must immediately instantiate an Article 17 Quality Management System for a system composed of components it does not fully own or control. Under point (c), modifying the intended purpose of a GPAI model by embedding it in a high-risk pipeline—for example, integrating a general-purpose chat model into a credit scoring workflow—triggers provider reclassification regardless of whether the model was independently classified as high-risk. Recital 88 explicitly contemplates this scenario: "Along the AI value chain multiple parties often supply AI systems, tools and services but also components or processes that are incorporated by the provider into the AI system."
Standards Gap	No harmonized standard provides guidance on when the assembly of components constitutes a substantial modification, when orchestration logic qualifies as a new system rather than a deployment configuration, or how the reclassified provider should structure the QMS for a system built from third-party components.
CRSA-1 Mechanism	Section 5 defines provider determination criteria for composed systems, including decision logic for classifying integration activities as substantial modification versus deployment configuration.

Article 25(2)–(3) — Cooperation Obligations

Text (paraphrased): Where the reclassification referred to in paragraph 1 takes place, the provider of the AI system as it was initially placed on the market or put into service shall closely cooperate with the new provider, make available the necessary technical documentation and capabilities, and provide the reasonably expected technical access and other assistance that are needed for the fulfilment of the obligations set out in this Regulation (paragraph 2). The initially placing provider shall not be required to provide information beyond what is necessary (paragraph 3).

Composition Relevance	Explicit
Compositional Implication	The "closely cooperate" obligation creates a bilateral duty between the original component provider and the new system provider (the enterprise deployer reclassified under Article 25(1)). For composed systems with multiple upstream component providers, this creates a web of cooperation obligations: the system provider must secure cooperation from every upstream provider whose component is integrated. The "reasonably expected" qualifier creates ambiguity: what technical access is "reasonably expected" when the upstream provider's model is a black-box GPAI system protected by trade secrets? The tension between the cooperation obligation and trade secret protection remains unresolved in the regulatory text. Article 78 provides a general confidentiality framework but does not resolve this specific tension for multi-vendor AI value chains.
Standards Gap	No harmonized standard defines the scope of "reasonably expected technical access" for GPAI model providers whose models are integrated into downstream composed systems, or the minimum technical documentation sufficient to satisfy cooperation obligations while protecting proprietary model internals.
CRSA-1 Mechanism	Appendix B provides Article 25(4) contract templates that define specific technical documentation deliverables, establishing a practical framework for what "closely cooperate" means in multi-vendor composed systems.

Article 25(4) — Mandatory Written Agreements

Text (paraphrased): The provider of a high-risk AI system and the third party that supplies an AI system, tools, services, components, or processes that are used or integrated in a high-risk AI system shall, by written agreement, specify the necessary information, capabilities, technical access, and other assistance, based on the generally acknowledged state of the art, in order to enable the provider of the high-risk AI system to fully comply with the obligations set out in this Regulation. The AI Office may develop and recommend voluntary model contractual terms.

Composition Relevance	Explicit
Compositional Implication	This is the operational backbone of the multi-vendor AI value chain. Article 25(4) mandates that every component supplier relationship in a composed system be governed by a written agreement specifying the information, capabilities, and technical access necessary for the system provider's compliance. For a composed system with four component suppliers, this means four bilateral agreements, each specifying documentation deliverables, technical access provisions, update notification mechanisms, and cooperation procedures. Without these agreements, the system provider cannot comply with: Article 9 (risk management requires information about component risk profiles), Article 10 (data governance requires information about component training data), Article 11 (technical documentation requires architectural information about third-party components), Article 12 (logging requires access to component-level event data), Article 13 (transparency requires information about component capabilities and limitations), or Article 15 (accuracy and robustness require component performance specifications). The AI Office is empowered to develop "voluntary model contractual terms" but has not published any as of March 2026. This absence leaves every company deploying a composed high-risk AI system without contractual infrastructure for the agreements the regulation requires.
Standards Gap	CRITICAL. No harmonized standard, Commission guideline, or AI Office publication provides model contract terms, minimum contractual provisions, or a template agreement for Article 25(4) compliance in multi-vendor AI systems.
CRSA-1 Mechanism	Appendix B of this specification provides the first published Article 25(4) contract template for composed AI systems, specifying minimum contractual provisions for information sharing, technical access, update notification, incident cooperation, and documentation deliverables.

2.12 Articles 26–27 — Deployer Obligations and Fundamental Rights Impact Assessment

Article 26 — Use in Accordance with Instructions

Text (paraphrased): Deployers shall use high-risk AI systems in accordance with the instructions of use accompanying the systems (paragraph 1), ensure that input data is relevant and sufficiently representative (paragraph 4), and monitor the operation of the high-risk AI system on the basis of the instructions of use and inform the provider or distributor of any serious incident or malfunction (paragraph 5).

Composition Relevance	Derived
Enforcement Date	2 August 2026
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	"Instructions of use" for a composed system must include compositional usage guidance—how the system should be deployed, configured, and monitored as an integrated whole. Deployers who modify the composed system—for example, by replacing a component model, connecting additional MCP tools, or altering routing thresholds—risk triggering the Article 25(1)(b) reclassification as a provider. The boundary between "using in accordance with instructions" and "making a substantial modification" is architecturally dependent: in a cascade system, changing the routing threshold is a configuration parameter; in the regulatory context, it may alter system behaviour sufficiently to affect Chapter III compliance. "Relevant and sufficiently representative" input data (paragraph 4) acquires compositional significance for RAG systems: the deployer's retrieval corpus constitutes operational input data whose relevance and representativeness directly affects system output quality. Deployer responsibility for input data quality in a RAG pipeline extends to retrieval corpus curation—an obligation the deployer may not recognise without compositional disclosure from the provider under Article 13.
Standards Gap	No harmonized standard provides guidance on instructions of use for composed systems, deployer input data obligations for RAG retrieval corpora, or the boundary between deployment configuration and substantial modification.
CRSA-1 Mechanism	Section 4 provides architecture-specific deployer guidance templates. Section 5 defines the configuration-versus-modification boundary for each composed system type.

Article 27 — Fundamental Rights Impact Assessment

Text (paraphrased): Prior to deploying a high-risk AI system referred to in Article 6(2) (Annex III systems), deployers that are bodies governed by public law, or are private entities providing public services, and deployers of systems referred to in Annex III points 5(b) and 5(c) (credit scoring and insurance risk assessment), shall perform an assessment of the impact of the use of the system on fundamental rights.

Composition Relevance	Derived
Compositional Implication	Evaluating fundamental rights impacts in a composed system requires compositional bias analysis. In a cascade routing system deployed by a government agency, if the small classifier disproportionately misclassifies speakers of certain dialects—routing them to a slower, lower-capability service pathway while routing native speakers to a frontier LLM—the compositional architecture itself creates structural discrimination. Neither model is individually discriminatory; the routing composition produces the discriminatory outcome. The fundamental rights impact assessment must evaluate the complete pipeline's impact on equality, non- discrimination, and access to services—not merely the performance of individual components.
Standards Gap	The AI Office guidelines on fundamental rights impact assessments do not address compositional bias analysis or sequential algorithmic discrimination arising from pipeline architecture. No harmonized standard or AI Office guideline provides a methodology for conducting fundamental rights impact assessments of composed AI systems, including analysis of sequential bias amplification, routing discrimination, or aggregation bias.
CRSA-1 Mechanism	Section 3 defines "Routing Discrimination" and "Aggregation Bias" as compositional risk categories. Section 4 provides architecture-specific fundamental rights assessment guidance including compositional bias detection methodologies.

2.13 Article 43 — Conformity Assessment

Article 43(1)–(3) — Assessment Procedures

Text (paraphrased): For Annex III high-risk systems, where the provider has applied harmonised standards and performed the relevant conformity assessment, the system shall follow the internal control procedure in Annex VI (paragraph 2). For remote biometric identification systems not covered by harmonised standards, conformity assessment shall follow Annex VII (third-party notified body assessment) (paragraph 1). For Annex I product-embedded systems, the sectoral conformity assessment applies with the AI Act requirements incorporated (paragraph 3).

Composition Relevance	Derived
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	The conformity assessment of a composed system faces a structural obstacle. The AI Act lacks a formalised mechanism for modular certification. A provider constructing a pipeline cannot aggregate the individual conformity evidence of component models to satisfy Article 43. The provider must generate system-level empirical evidence proving that the integrated whole is compliant, evaluating the semantic and technical boundaries where models connect. For Annex III systems applying the internal control procedure (Annex VI), the provider self-certifies compliance of the entire composed system—requiring the provider to possess sufficient technical understanding of all components to make this certification. For biometric systems or Annex I product-embedded systems requiring third-party assessment (Annex VII), the notified body must evaluate the composed system as a whole—requiring assessment methodologies that no notified body has published. As of March 2026, no AI Act-specific notified bodies have been formally designated. The likely candidates—TÜV SÜD, TÜV Rheinland, Bureau Veritas, DNV, SGS, DEKRA—are developing AI assessment capabilities but none has published a methodology for multi-model AI system evaluation.
Standards Gap	CRITICAL. prEN 18285 (Conformity Assessment Framework) provides no methodology for conformity assessment of systems composed of multiple AI components from different providers. No notified body has published an assessment methodology for multi-model systems.
CRSA-1 Mechanism	Section 5 provides a complete conformity assessment methodology for composed systems, covering both internal control (Annex VI) and third-party assessment (Annex VII) procedures adapted for multi-model architectures.

Article 43(4) — Predetermined Changes and Re-Assessment

Text (paraphrased): Where a modification of a high-risk AI system takes place that has been pre-determined by the provider and assessed at the moment of the initial conformity assessment, that modification shall not require a new conformity assessment. In all other cases, where the modification may affect the compliance of the high-risk AI system with the requirements or may modify the system's intended purpose, a new conformity assessment shall be required.

Composition Relevance	Explicit
Compositional Implication	This provision is simultaneously the primary mitigation and the primary trap for composed system providers. If the provider pre-documents foreseeable component changes (model version upgrades, retrieval corpus updates, tool additions) as predetermined changes in the initial conformity assessment, those changes can proceed without re-assessment. If the provider fails to pre-document a change, executing it triggers full re-assessment. The incentive structure is clear: providers of composed systems should pre-document every conceivable component change and its acceptable parametric range. However, this creates tension with assessment specificity—an assessment that pre-approves all possible changes provides weaker compliance evidence than an assessment of a specific, bounded system. The balance between pre-determination breadth and assessment rigour is an engineering discipline that no standard defines. For MCP-based multi-agent systems, the pre-determination challenge is acute: the system's capabilities change when new tools are connected. Pre-documenting all possible tool connections may be infeasible if the tool ecosystem is open and externally governed.
Standards Gap	No harmonized standard provides a methodology for structuring pre-determined change documentation for composed systems or for defining acceptable parametric ranges for component modifications.
CRSA-1 Mechanism	Section 5 provides a pre-determination framework for composed systems, including parametric boundary specification for each architecture type and guidance on balancing pre-determination breadth with assessment rigour.

2.14 Articles 51–56 — General-Purpose AI Models and Systemic Risk

Article 51 — Classification and Systemic Risk Designation

Text (paraphrased): A GPAI model shall be classified as a GPAI model with systemic risk if it has high-impact capabilities, which is presumed when the cumulative amount of computation used for its training measured in floating-point operations exceeds 10^{25} (paragraph 2).

Composition Relevance	Derived
Enforcement Date	2 August 2025 (GPAI provisions already applicable)
Compositional Implication	The systemic risk classification uses a training compute threshold applied to individual models. The AI Act provides no mechanism to assess whether composing multiple non-systemic GPAI models creates systemic risk through emergent aggregate capabilities. A multi-agent framework deploying several specialised models—each trained below the 10^{25} FLOPs threshold—could theoretically generate autonomous capabilities that equal or exceed those of a systemic risk model. Annex XIII criteria for systemic risk assessment include reach, scalability, access to tools, and degree of autonomy—all properties that composition can enhance without increasing any individual model's training compute. The AI Office has not clarified whether the FLOPs threshold or the systemic risk designation can be applied holistically to the aggregate inference compute and combined logic of multi-agent systems. This ambiguity affects both GPAI providers (whose models may collectively create systemic risk when composed) and system providers (who may be unknowingly operating a system with systemic-risk-level capabilities).
Standards Gap	The General-Purpose AI Code of Practice addresses individual GPAI model obligations. No standard or guidance addresses the emergent systemic risk assessment of composed GPAI model deployments.
CRSA-1 Mechanism	Section 3 defines "Capability Drift" as a compositional risk category that captures unbounded capability expansion through composition. Section 4 (MCP Multi-Agent profile) addresses the systemic risk assessment implications of dynamic capability extension.

Article 53 — Obligations of GPAI Model Providers

Text (paraphrased): Providers of GPAI models shall draw up and keep up to date the technical documentation of the model, including its training and testing process and the results of its evaluation, and make available to providers of AI systems who intend to integrate the GPAI model into their systems the information and documentation that is necessary to enable those providers to comply with their obligations under this Regulation.

Composition Relevance	Explicit
Compositional Implication	Article 53 creates the upstream documentation obligation that enables downstream compositional safety compliance. The GPAI provider must provide "information and documentation necessary to enable" downstream system providers to comply with Articles 9, 10, 11, 12, 13, and 15. For composed systems, the adequacy of this upstream documentation for compositional safety assessment is the critical question. Annex XII specifies the technical documentation GPAI providers must maintain, including training methodologies, data curation practices, capability evaluations, and known limitations. The downstream system provider integrating this GPAI model into a composed pipeline needs this information to assess compositional risk (Article 9), document third-party components (Annex IV 2(a)), and specify system-level accuracy (Article 15(2)). If the upstream documentation is insufficient for compositional assessment, the system provider faces a compliance gap that no contractual provision under Article 25(4) can fully resolve—the upstream provider cannot disclose what they have not evaluated.
Standards Gap	No harmonized standard defines the minimum GPAI model documentation sufficient for downstream compositional safety assessment, or specifies what additional information beyond Annex XII a system provider requires to evaluate compositional risk.
CRSA-1 Mechanism	Section 5 specifies minimum upstream documentation requirements for GPAI components integrated into composed high-risk systems, extending beyond Annex XII to address compositional assessment needs.

2.15 Articles 61 and 72 — Post-Market Monitoring

Article 61(1)–(3) and Article 72 — Monitoring System

Text (paraphrased): Providers shall establish and document a post-market monitoring system in a manner that is proportionate to the nature of the AI technologies and the risks of the high-risk AI system (Article 61(1)). The post-market monitoring system shall actively and systematically collect, document, and analyse relevant data which may be provided by deployers or which may be collected through other sources on the performance of high-risk AI systems throughout their lifetime (Article 61(3)). Article 72 requires a post-market monitoring plan as part of the technical documentation referenced in Article 11.

Composition Relevance	Implicit
Enforcement Date	2 August 2026 (Annex III); 2 August 2027 (Annex I)
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Post-market monitoring of a composed system requires tracking operational drift not merely in the final user-facing output but in the internal component interactions and intermediate processing stages. In an ensemble system, monitoring must detect when the voting consensus mechanism shifts—when one component model begins dominating decisions or when aggregation weights drift from their assessed baseline. In a RAG pipeline, monitoring must detect when retrieval relevance degrades or when the semantic alignment between retrieval and generation components deteriorates. The "collected through other sources" language in Article 61(3) is significant for composed systems: the system provider may need to collect performance data from upstream component providers to monitor compositional health. If an upstream provider silently updates a model API, the downstream system provider's post-market monitoring must detect the change through its effects on system-level performance—requiring compositional monitoring infrastructure that observes inter-component behaviour, not merely endpoint outputs. The Digital Omnibus proposes removing the mandatory post-market monitoring plan template, but the substantive monitoring obligation remains.
Standards Gap	No harmonized standard provides a post-market monitoring framework for composed AI systems, including compositional drift detection, inter-component performance correlation, or cross-provider telemetry integration.
CRSA-1 Mechanism	Section 4 provides architecture-specific post-market monitoring requirements, including compositional health metrics, drift detection triggers, and inter-provider monitoring coordination mechanisms.

2.16 Article 86 — Right to Explanation for Individual Decision-Making

Article 86(1) — Explanation of AI-Assisted Decisions

Text (paraphrased): Any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III, and which produces legal effects or similarly significantly affects that person, shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.

Composition Relevance	Derived
Enforcement Date	2 August 2026
Penalty Tier	**€15M or 3% global turnover**
Compositional Implication	Providing a "clear and meaningful explanation" of a decision produced by a composed system requires compositional explainability—the ability to trace the decision through the pipeline and identify the contribution of each component to the final output. In an ensemble employment screening system, a rejected applicant's right to explanation cannot be satisfied by disclosing that "the AI system rejected the application." The explanation must convey how the component models contributed to the aggregate decision: which models flagged concerns, how the voting mechanism weighted those flags, and what the decisive factors were at the system level. In a RAG pipeline producing a credit decision, the explanation must convey the role of retrieved context (what information was retrieved and why), the generation process (how the model synthesised the retrieved information), and the decision logic (how the output was translated into the credit decision). Current explainability standards, including ISO/IEC 24028, do not specify interpretability metrics for composed systems—such as Shapley value decomposition across pipeline stages or causal attribution across multi-model decision chains.
Standards Gap	No harmonized standard provides a methodology for generating explanations of decisions produced by composed AI systems, including cross-component causal attribution, ensemble vote decomposition, or pipeline-stage contribution analysis.
CRSA-1 Mechanism	Section 4 provides architecture-specific explainability requirements, including minimum explanation elements for each composed system type and guidance on compositional causal attribution.

2.17 Obligation Mapping Summary

The following table summarises the composition relevance of each mapped obligation. Of the twenty-seven distinct obligations analysed, seven contain explicit compositional language, twelve require implicit compositional analysis, and eight derive compositional requirements from system-level application.

Provision	Comp. Type	Enforcement	Penalty Tier	Standards Gap
Art. 3(1) — AI System Definition	Derived	—	—	ISO/IEC 22989
Art. 3(3) — Provider Definition	Explicit	—	—	prEN 18286
Art. 3(23) — Substantial Modification	Explicit	—	—	None published
Art. 3(63) — GPAI Model Definition	Explicit	—	—	GPAI CoP
Art. 6(1)–(2) — Classification	Implicit	Aug 2026/27	€15M / 3%	None published
Art. 6(3) — Derogation	Derived	Aug 2026/27	€15M / 3%	None published
Art. 8(1)–(2) — Compliance Framework	Derived	Aug 2026/27	€15M / 3%	prEN 18286
Art. 9(1)–(2) — Risk Management	Implicit	Aug 2026/27	€15M / 3%	prEN 18228
Art. 9(4) — Interaction Effects	Explicit	Aug 2026/27	€15M / 3%	prEN 18228
Art. 9(5)–(8) — Testing & Misuse	Derived	Aug 2026/27	€15M / 3%	prEN 18229-2
Art. 10(1)–(2) — Data Governance	Implicit	Aug 2026/27	€15M / 3%	prEN 18284
Art. 10(2)(f) — Bias Examination	Derived	Aug 2026/27	€15M / 3%	None published
Art. 10(5) — Operational Input Data	Derived	Aug 2026/27	€15M / 3%	prEN 18284
Art. 11 / Annex IV 2(a) — Third-Party	Explicit	Aug 2026/27	€15M / 3%	None published
Annex IV 2(c) — Architecture	Explicit	Aug 2026/27	€15M / 3%	Critical
Annex IV 1(b) — System Interactions	Explicit	Aug 2026/27	€15M / 3%	None published
Annex IV 2(f) — Predetermined Changes	Explicit	Aug 2026/27	€15M / 3%	None published
Art. 12(1)–(2) — Logging	Implicit	Aug 2026/27	€15M / 3%	prEN 18229-1
Art. 12(2) — Risk Identification	Derived	Aug 2026/27	€15M / 3%	None published
Art. 13(1) — Transparency	Implicit	Aug 2026/27	€15M / 3%	prEN 18229-1
Art. 13(3)(b) — Capabilities	Derived	Aug 2026/27	€15M / 3%	None published
Art. 14(1)–(2) — Oversight Design	Implicit	Aug 2026/27	€15M / 3%	prEN 18229-1
Art. 14(4)(a) — Understanding	Derived	Aug 2026/27	€15M / 3%	None published

Provision	Comp. Type	Enforcement	Penalty Tier	Standards Gap
Art. 14(4)(d)–(e) — Override / Stop	Implicit	Aug 2026/27	€15M / 3%	CRITICAL
Art. 15(1) — Accuracy / Robustness	Derived	Aug 2026/27	€15M / 3%	prEN 18229-2
Art. 15(2) — Declared Metrics	Derived	Aug 2026/27	€15M / 3%	CRITICAL
Art. 15(3) — Error Resilience	Explicit	Aug 2026/27	€15M / 3%	ISO/IEC 24029
Art. 15(4)–(5) — Cybersecurity	Derived	Aug 2026/27	€15M / 3%	prEN 18282
Art. 25(1) — Provider Reclass.	Explicit	Aug 2026	€15M / 3%	None published
Art. 25(2)–(3) — Cooperation	Explicit	Aug 2026	€15M / 3%	None published
Art. 25(4) — Written Agreements	Explicit	Aug 2026	€15M / 3%	CRITICAL
Art. 26 — Deployer Use	Derived	Aug 2026	€15M / 3%	None published
Art. 27 — FRIA	Derived	Aug 2026	€15M / 3%	None published
Art. 43(1)–(3) — Conformity Assess.	Derived	Aug 2026/27	€15M / 3%	CRITICAL
Art. 43(4) — Predetermined Changes	Explicit	Aug 2026/27	€15M / 3%	None published
Art. 51 — Systemic Risk	Derived	Aug 2025	Art. 101 regime	GPAI CoP
Art. 53 — GPAI Documentation	Explicit	Aug 2025	Art. 101 regime	Annex XII
Art. 61 / 72 — Post-Market	Implicit	Aug 2026/27	€15M / 3%	None published
Art. 86 — Right to Explanation	Derived	Aug 2026	€15M / 3%	ISO/IEC 24028

Five obligations are flagged as CRITICAL—indicating that the compositional gap is absolute, with no partial coverage from any existing or in-development standard:

1. **Annex IV, point 2(c)**—system architecture documentation for composed systems (mandate without method).
2. **Article 14(4)(d)–(e)**—safe interruption and state rollback in multi-agent systems (mandate without mechanism).
3. **Article 15(2)**—declared accuracy metrics for composed systems (specification obligation without specification methodology).
4. **Article 25(4)**—written agreements for multi-vendor AI value chains (contractual mandate without model terms).

5. **Article 43(1)–(3)**—conformity assessment of composed systems (assessment obligation without assessment methodology).

These five critical gaps define the minimum scope that any compositional safety specification must address to provide compliance infrastructure for composed AI systems under the EU AI Act. Sections 3 through 5 and Appendix B of this specification are structured to address each.

3 Compositional Risk Taxonomy

3.1 Purpose and Regulatory Basis

Article 9(2) of Regulation (EU) 2024/1689 requires the identification and analysis of "known and reasonably foreseeable risks" that a high-risk AI system can pose. Article 9(4) further requires that risk management measures give "due consideration to the effects and possible interaction resulting from the combined application of the requirements." For composed AI systems, these obligations demand a risk classification framework that captures failure modes arising specifically from the interaction between components— failure modes that are absent from any individual component's risk profile.

No harmonized standard under development within CEN-CENELEC JTC 21 provides such a framework. prEN 18228 (AI Risk Management) addresses risk at the organisational and single-system level. ISO/IEC 23894 extends ISO 31000 to AI but assumes identifiable, bounded systems. Neither provides a taxonomy of risks that emerge from composition.

This section establishes the **CRSA-1 Compositional Risk Taxonomy**: eight categories of risk specific to composed AI systems, each mapped to the AI Act obligations it affects, with detection methodology and mitigation patterns defined at the architectural level. This taxonomy provides the vocabulary and classification structure that Article 9 requires but that no existing standard delivers.

3.2 Taxonomy Architecture

The eight compositional risk categories are organised along two dimensions: the *origin* of the risk (where in the composition the failure initiates) and the *propagation mode* (how the failure manifests across the system). The taxonomy is exhaustive with respect to the five composed architecture types defined in Section 1.2 and the thirty-eight obligations mapped in Section 2.

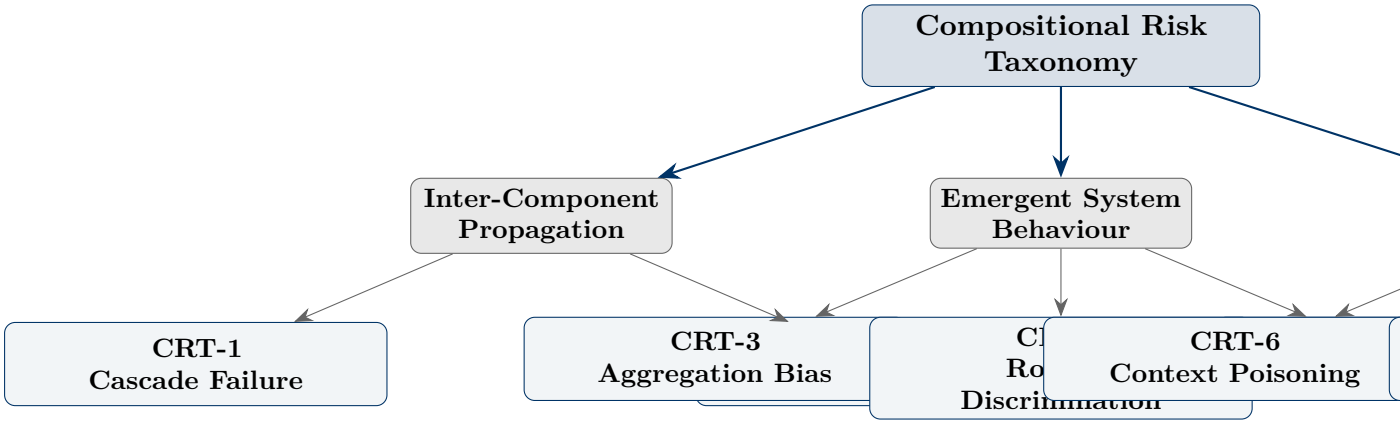

Figure 1: CRSA-1 Compositional Risk Taxonomy. Eight risk categories organised by origin: inter-component propagation, emergent system behaviour, and boundary/interface failures.

3.3 Risk Category Definitions

3.3.1 CRT-1: Cascade Failure

Definition

A cascade failure occurs when an error, fault, or inconsistency generated by one component in a composed AI system propagates through downstream components, amplifying in severity or altering in character at each stage. The downstream components process the erroneous input as valid, producing outputs that are systematically distorted by the upstream fault. The distinguishing characteristic of cascade failure is that the error is *invisible at each individual component boundary*—each component performs its function correctly given its input, but the chain of correct local operations produces an incorrect global output.

Origin: Inter-component propagation.
Affected Obligations:

- Article 9(1)–(2) — risk management must identify cascade pathways as foreseeable risks.
- Article 12(2) — logging must enable identification of the originating component and propagation path.
- Article 15(1) — system-level accuracy degrades through cascading error even when component-level accuracy is maintained.
- Article 15(3) — resilience against errors "within the system" explicitly encompasses cascade propagation.

Primary Architectures: RAG pipelines (retrieval error propagating to generation), agentic systems (planner error propagating through executor), cascade/routing systems (misrouted input processed by inappropriate model).

Detection Methodology: Cascade failure is detected by monitoring the divergence between expected and observed output distributions at each pipeline stage. When Component n's output distribution deviates beyond its assessed baseline, and Component $n{+}1$'s output distribution shows correlated deviation, a cascade pathway is indicated. Cross-component correlation monitoring—tracking the statistical dependency between sequential component outputs—provides the detection signal.

Mitigation Pattern: Cascade failure mitigation requires inter-component validation gates: verification points between pipeline stages that assess the plausibility of intermediate outputs before passing them to the next component. Validation gates operate on the output distribution of each stage, flagging outputs that fall outside the assessed operational envelope. Specific threshold parameters and gate implementation specifications are architecture-dependent.

EU Compliance Series — Forthcoming

Detailed threshold calibration methodologies for cascade failure detection, including statistical divergence measures, correlation window specifications, and validation gate sensitivity parameters for each architecture type, will be provided in sector-specific editions of the CRSA-1 EU Compliance Series.

3.3.2 CRT-2: Semantic Drift

Definition

Semantic drift occurs when the meaning or representational content of data is altered as it passes between components that operate in different embedding spaces, use different tokenisation schemes, or maintain different internal representations of the same concepts. Unlike cascade failure, where the propagated content is erroneous, semantic drift involves content that is *transformed* across a model boundary—the information is not incorrect at any single point, but its meaning shifts in transit. The downstream component processes a semantically different input than the upstream component intended to produce.

Origin: Inter-component propagation.
Affected Obligations:

- Article 10(2) — data governance must ensure representational consistency across the data pipeline.
- Article 11 / Annex IV 2(c) — documentation must characterise the semantic boundaries between components.
- Article 15(1) — system-level accuracy is degraded by semantic drift even when component-level accuracy is preserved.
- Article 15(2) — declared accuracy metrics must account for semantic drift as a source of system-level error.

Primary Architectures: RAG pipelines (embedding model and generation model operating in different representational spaces), ensemble systems (component models interpreting the same input differently), MCP-based systems (tool outputs interpreted by models trained on different data distributions).

Detection Methodology: Semantic drift is detected through representational alignment monitoring: measuring the cosine similarity or other distance metrics between the intended semantic content of an upstream output and the downstream component's internal representation of that content. Persistent reduction in alignment scores indicates semantic drift. For RAG systems specifically, monitoring the relevance score distribution between retrieved documents and generated outputs provides a proxy signal.

Mitigation Pattern: Semantic drift mitigation requires representational calibration between connected components: ensuring that the output space of Component n is aligned with the input expectations of Component $n{+}1$. This may involve shared embedding spaces, calibration layers, or explicit semantic verification at component boundaries.

EU Compliance Series — Forthcoming

Representational alignment benchmarks, calibration layer specifications, and semantic verification protocols for each architecture type will be provided in subsequent editions.

3.3.3 CRT-3: Aggregation Bias

Definition

Aggregation bias occurs when individually compliant AI components—each satisfying bias and fairness requirements independently—produce a discriminatory outcome when their outputs are combined through voting, averaging, weighting, or other aggregation mechanisms. The bias is an emergent property of the aggregation architecture, not a property of any individual component. Aggregation bias can amplify minor component biases, introduce new bias patterns absent from any component, or mask component biases in ways that evade standard single-model fairness testing.

Origin: Emergent system behaviour.
Affected Obligations:

- Article 9(1)–(2) — risk management must assess aggregation bias as a foreseeable compositional risk.
- Article 10(2)(f) — bias examination must extend to the aggregation mechanism, not merely individual training datasets.
- Article 27 — fundamental rights impact assessment must evaluate the discriminatory effects of the aggregation architecture.
- Article 86 — right to explanation must account for how the aggregation mechanism contributed to the decision.

Primary Architectures: Ensemble systems (voting aggregation producing emergent bias), cascade/routing systems (routing logic creating differential service quality).

Detection Methodology: Aggregation bias is detected by comparing fairness metrics (demographic parity, equalised odds, calibration) at both the component level and the system level. When component-level metrics satisfy fairness thresholds but system-level metrics do not, the discrepancy indicates aggregation bias. The detection requires simultaneous monitoring of both levels—a compositional fairness analysis that no current standard mandates.

Mitigation Pattern: Aggregation bias mitigation requires fairness-aware aggregation design: structuring the voting, weighting, or consensus mechanism to preserve component-level fairness properties at the system level. This may include constrained aggregation algorithms, fairness regularisation in weighting schemes, or post-aggregation bias correction.

3.3.4 CRT-4: Routing Discrimination

Definition

Routing discrimination occurs when a classifier, router, or threshold mechanism within a composed system directs inputs to different processing pathways in a manner that correlates with protected characteristics. The different pathways provide different levels of service quality, accuracy, or capability, resulting in systematically differential treatment of demographic groups. The discrimination arises from the routing architecture, not from any individual model's bias—the router may be statistically neutral across protected groups on standard fairness metrics, yet produce discriminatory outcomes through interaction with downstream pathway quality differentials.

Origin: Emergent system behaviour.
Affected Obligations:

- Article 9(4) — the interaction between the routing mechanism and downstream model capabilities is precisely the "effect and possible interaction" that Article 9(4) requires risk management to assess.

- Article 27 — fundamental rights impact assessment must evaluate whether routing creates differential access to service quality.
- Article 15(1) — system-level accuracy varies by routing pathway, creating population-dependent accuracy that must be declared.
- Article 86 — affected persons routed to lower-quality pathways have the right to explanation of the routing decision.

Primary Architectures: Cascade/routing systems (classifier directing to small versus large model), agentic systems (planner selecting different tool chains for different input profiles).

Detection Methodology: Routing discrimination is detected by analysing the correlation between routing decisions and protected characteristics, disaggregated by downstream pathway. The analysis must evaluate not merely whether the router is fair in isolation, but whether the combined effect of routing decision and pathway quality differential produces discriminatory outcomes. This requires joint analysis of the router's decision boundary and the performance differential between downstream pathways.

Mitigation Pattern: Routing discrimination mitigation requires either equalising downstream pathway quality (so that routing decisions do not create service differentials) or applying fairness-constrained routing (ensuring that protected groups are not disproportionately directed to lower-quality pathways). The choice between these approaches involves cost-accuracy trade-offs that are deployment- specific.

3.3.5 CRT-5: Capability Drift

Definition

Capability drift occurs when the functional capabilities of a composed AI system change after deployment through the addition, removal, or modification of connected tools, data sources, or component models—without a corresponding update to the system's conformity assessment, documentation, or risk management. The system that operates is no longer the system that was assessed. Capability drift is distinct from performance degradation: the system does not become worse at its assessed functions, but acquires new functions or loses existing ones in ways that alter its risk profile.

Origin: Emergent system behaviour.
Affected Obligations:

- Article 3(23) — capability drift may constitute a substantial modification if the change was not predetermined and affects compliance.
- Article 11 / Annex IV 1(b) — documentation of system interactions becomes stale when capabilities change.
- Article 13(3)(b) — declared capabilities and limitations no longer reflect the operational system.
- Article 43(4) — unless the capability change was predetermined, it triggers re-assessment.
- Article 51 — aggregate capabilities may cross the systemic risk threshold through composition without any individual model change.

Primary Architectures: MCP-based multi-agent systems (tool connections altering system capabilities), agentic systems (dynamic tool selection expanding operational scope).

Detection Methodology: Capability drift is detected through continuous capability enumeration: maintaining a runtime inventory of connected tools, active data sources, and available model endpoints, and comparing the operational inventory against the assessed inventory documented in the conformity assessment. Any discrepancy indicates capability drift.

Mitigation Pattern: Capability drift mitigation requires capability governance: a control mechanism that prevents the composed system from activating tools, data sources, or model

connections that were not included in the conformity assessment, unless those additions were pre-documented as predetermined changes under Annex IV 2(f). For MCP-based systems, this means implementing a tool allowlist assessed at conformity assessment time, with additions requiring explicit compliance review.

3.3.6 CRT-6: Context Poisoning

Definition

Context poisoning occurs when adversarial content is introduced into the data environment from which a composed AI system retrieves or receives context, causing downstream components to act on malicious instructions or corrupted information. The adversarial content is not directed at any individual component's input interface; it is placed in the shared data environment and activated when a component retrieves or processes it. This is the compositional form of indirect prompt injection: the attack surface exists because the system composes retrieval with execution.

Origin: Boundary and interface.
Affected Obligations:

- Article 15(4)–(5) — cybersecurity must address adversarial exploitation of the compositional attack surface.
- Article 9(8) — foreseeable misuse includes adversarial manipulation of retrieval corpora and tool outputs.
- Article 12(1) — logging must capture the provenance of context consumed by each component to enable forensic reconstruction of poisoning incidents.
- Article 25(4) — written agreements must address security responsibilities for shared data environments.

Primary Architectures: RAG pipelines (retrieval corpus poisoning), MCP-based systems (adversarial content in connected data sources), agentic systems (tool outputs containing injected instructions).

Detection Methodology: Context poisoning is detected through input provenance verification: validating the integrity and trustworthiness of all context consumed by each component before processing. This includes retrieval corpus integrity monitoring (detecting unauthorised modifications to indexed documents), tool output sanitisation (detecting instruction-like content in data responses), and cross-component input anomaly detection (identifying context that deviates from expected distributions).

Mitigation Pattern: Context poisoning mitigation requires compositional sandboxing: ensuring that context retrieved or received from external sources is processed in an isolated evaluation environment before being passed to execution-capable components. The separation between retrieval and execution—where retrieved content is treated as untrusted data rather than executable instruction—is the architectural control.

3.3.7 CRT-7: State Corruption

Definition

State corruption occurs when the interruption, failure, or timeout of one component in a composed AI system leaves shared state—databases, context windows, external system records, or inter-agent memory—in an inconsistent or partially updated condition. Downstream components or subsequent operations inherit the corrupted state, producing outputs based on incomplete or contradictory information. State corruption is the compositional consequence of Article 14(4)(e)'s "stop button" requirement: halting a composed system mid-operation creates the conditions for state corruption unless the architecture provides transactional guarantees across component boundaries.

Origin: Boundary and interface.
Affected Obligations:

- Article 14(4)(d)–(e) — safe interruption requires preventing state corruption upon halt.
- Article 15(3) — resilience against faults "within the system" includes resilience against state corruption from component failure.
- Article 12(1) — logging must capture the state of all components at the point of interruption or failure.
- Article 61 — post-market monitoring must detect state corruption incidents and their downstream effects.

Primary Architectures: Agentic systems (tool-calling agent interrupted mid-write to external database), MCP-based systems (multi-agent process halted with partial state updates across connected systems), ensemble systems (voting interrupted mid-aggregation).

Detection Methodology: State corruption is detected through state consistency verification: comparing the expected state (derived from the complete operation specification) with the actual state (observed across all shared state repositories) after any interruption, timeout, or component failure. Inconsistencies indicate state corruption.

Mitigation Pattern: State corruption mitigation requires transactional composition: ensuring that multi-component operations either complete fully or roll back entirely across all affected state repositories. This is the compositional equivalent of database transaction guarantees (ACID properties) applied to multi-agent AI workflows. The architecture must support atomic operations across component boundaries, with rollback capability for interrupted processes.

3.3.8 CRT-8: Compositional Opacity

Definition

Compositional opacity occurs when the interaction between components in a composed AI system produces decision processes that cannot be meaningfully explained, attributed, or decomposed into component contributions. Each individual component may be interpretable in isolation, but the composed decision pathway—the sequence of component interactions that produced the output—resists explanation because the causal relationships between component contributions are non-linear, context-dependent, or emergent. Compositional opacity is distinct from individual model opacity: a pipeline of individually interpretable models may produce compositionally opaque decisions.

Origin: Boundary and interface.
Affected Obligations:

- Article 13(1) — transparency sufficient to "interpret the system's output" requires overcoming compositional opacity.

- Article 14(4)(a) — human overseers must "fully understand the relevant capacities and limitations," which compositional opacity impedes.
- Article 86 — the right to a "clear and meaningful explanation" of a composed decision requires compositional explainability methods that decompose the decision across pipeline stages.

Primary Architectures: All composed architectures, with severity proportional to pipeline depth and the number of non-linear interaction points. Ensemble systems with non-linear voting aggregation and agentic systems with dynamic tool-calling chains present the highest compositional opacity.

Detection Methodology: Compositional opacity is assessed through explainability coverage analysis: measuring the proportion of the composed decision pathway for which causal attribution can be established. When significant portions of the decision pathway cannot be attributed to specific component contributions or explained through input-output relationships, the system exhibits compositional opacity that exceeds the transparency threshold of Article 13.

Mitigation Pattern: Compositional opacity mitigation requires structured interpretability architecture: designing the composed system so that each component boundary produces an interpretable intermediate output, and the aggregation of these intermediate outputs into the final decision follows an attributable logic. This may include pipeline stage contribution tracking, ensemble vote decomposition, or agent decision chain logging that enables post-hoc causal reconstruction.

EU Compliance Series — Forthcoming

Explainability coverage metrics, causal attribution methodologies for each architecture type, and minimum interpretability requirements for Article 86 compliance in composed systems will be provided in subsequent editions of the CRSA-1 EU Compliance Series.

3.4 Cross-Reference: Risk Categories to Obligations

Table 2 maps each compositional risk category to its primary affected obligations and the architecture types most susceptible to each risk. This mapping provides the foundation for the architecture-specific compliance profiles in Section 4.

Risk Category	Primary Obligations	Primary Architectures	Detection Signal
CRT-1 Cascade Failure	Art. 9, 12, 15(1), 15(3)	RAG, Agentic, Cascade	Cross-component output correlation divergence
CRT-2 Semantic Drift	Art. 10, 11, 15(1), 15(2)	RAG, Ensemble, MCP	Representational alignment score degradation
CRT-3 Aggregation Bias	Art. 9, 10(2)(f), 27, 86	Ensemble, Cascade	Component vs. system fairness metric divergence
CRT-4 Routing Discrim.	Art. 9(4), 15(1), 27, 86	Cascade, Agentic	Protected characteristic correlation with routing

Risk Category	Primary Obligations	Primary Architectures	Detection Signal
CRT-5 Capability Drift	Art. 3(23), 11, 13, 43(4), 51	MCP, Agentic	Operational vs. assessed capability inventory mismatch
CRT-6 Context Poisoning	Art. 9(8), 12, 15(4)–(5), 25(4)	RAG, MCP, Agentic	Input provenance anomaly; instruction-like data content
CRT-7 State Corruption	Art. 12, 14(4), 15(3), 61	Agentic, MCP, Ensemble	Post-interruption state consistency verification failure
CRT-8 Comp. Opacity	Art. 13, 14(4)(a), 86	All (severity varies)	Explainability coverage below attribution threshold

Table 2: Compositional Risk Taxonomy cross-reference. Each risk category mapped to primary AI Act obligations, most susceptible architectures, and detection signal.

3.5 Taxonomy Completeness and Extensibility

This taxonomy is designed to be exhaustive with respect to the compositional failure modes identifiable across the five architecture types addressed in this specification as of March 2026. The taxonomy is structured for extensibility: as composed AI architectures evolve and new interaction patterns emerge, additional risk categories may be defined following the same structure (definition, origin classification, affected obligations, detection methodology, mitigation pattern).

New risk categories **SHALL** be assigned sequential CRT identifiers (CRT-9 and beyond) and **SHALL** be mapped to the obligation framework established in Section 2. Subsequent editions of the CRSA-1 EU Compliance Series will incorporate taxonomy extensions as warranted by architectural developments and regulatory guidance.

The taxonomy identifiers (CRT-1 through CRT-8) are referenced throughout the remainder of this specification. Section 4 maps each architecture profile to its applicable CRT categories. Section 5 structures conformity assessment evidence requirements around the CRT framework. Appendix A cross-references CRT categories with CRSA-1 protocol elements and AI Act provisions.

4 Architecture-Specific Compliance Profiles

4.1 Purpose and Structure

The obligation mapping in Section 2 establishes *what* the AI Act requires for composed systems. The risk taxonomy in Section 3 establishes *what can go wrong*. This section establishes *how* specific composed architectures must be designed, documented, monitored, and assessed to satisfy those requirements and manage those risks.

Each profile addresses the five dominant composed AI architecture types identified in Section 1.2. For each architecture, the profile provides:

1. An architecture reference diagram identifying components, data flows, and compliance-critical interaction points.

2. The applicable compositional risk categories (CRT-1 through CRT-8) ranked by severity for that architecture.
3. Article-by-article compliance requirements specific to that architecture type, addressing the critical gaps identified in Section 2.
4. Logging architecture requirements satisfying Article 12.
5. Human oversight intervention architecture satisfying Article 14.
6. Accuracy and robustness specification methodology satisfying Article 15.

These profiles are designed to be directly applicable: a provider deploying a RAG pipeline for credit scoring can use Profile 4.2 as the structural basis for their Annex IV technical documentation, Article 9 risk management plan, and Article 12 logging specification.

4.2 Profile 1: RAG Pipeline in Regulated Decision-Making

4.2.1 Architecture Reference

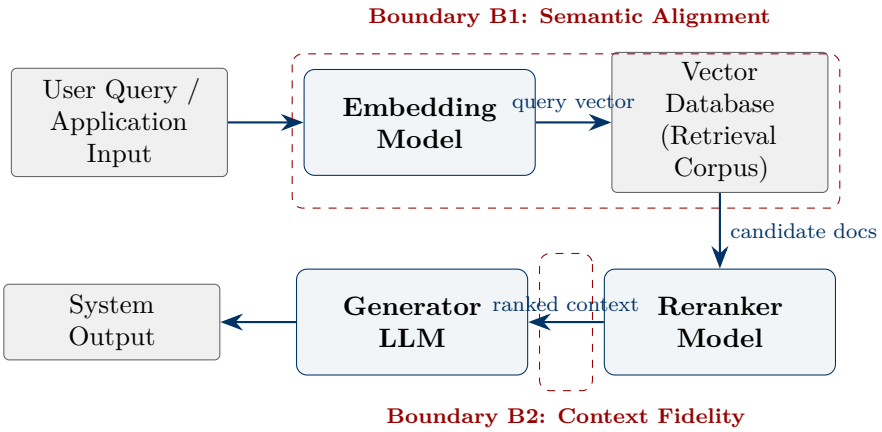

Figure 2: RAG Pipeline architecture reference. Two compliance-critical boundaries identified: B1 (semantic alignment between embedding and retrieval) and B2 (context fidelity between reranking and generation).

4.2.2 Applicable Risk Categories

Risk Category	Severity	RAG-Specific Manifestation
CRT-1 Cascade Failure	High	Retriever returns irrelevant documents; generator synthesises them into a confident but incorrect output. Neither component is individually faulty.
CRT-2 Semantic Drift	High	Embedding model and generator operate in different representational spaces. The "meaning" of a query shifts between retrieval and generation.
CRT-6 Context Poisoning	High	Adversarial content injected into the retrieval corpus is retrieved and acted upon by the generator as trusted context.
CRT-8 Comp. Opacity	Medium	The causal pathway from retrieved document to generated output is non-transparent: the deployer cannot determine which retrieved documents influenced which portions of the output.
CRT-3 Aggregation Bias	Low–Medium	When multiple retrieved documents reflect different demographic perspectives, the generator's synthesis may disproportionately weight certain perspectives.

4.2.3 Compliance Requirements by Article

Article 9 — Risk Management. The risk management system for a RAG pipeline **SHALL** identify and assess the following compositional risk pathways:

1. **Retrieval relevance failure:** the probability that the retriever returns documents that are topically related but factually inapplicable to the query, and the downstream impact on generator output accuracy.
2. **Retrieval corpus staleness:** the risk that the corpus contains outdated information that the generator presents as current, particularly in domains with rapid regulatory or factual change.
3. **Semantic misalignment:** the risk that the embedding model's representation of the query diverges from the generator's interpretation of retrieved context, producing outputs that are internally coherent but semantically disconnected from the user's intent.
4. **Corpus poisoning:** the risk that adversarial or erroneous content in the retrieval corpus is surfaced and acted upon by the generator (CRT-6).

Risk management measures **SHALL** give due consideration to the interaction between retrieval quality and generation accuracy as required by Article 9(4). Optimising retrieval recall (returning more candidate documents) may degrade generation precision (increasing the probability of the generator incorporating irrelevant context).

Article 11 / Annex IV 2(c) — Architecture Documentation. The technical documentation for a RAG pipeline **SHALL** include, at minimum:

1. The embedding model specification, including embedding dimensions, training domain, and representational coverage.
2. The vector database configuration, including indexing algorithm, similarity metric, and retrieval parameters (top-k, similarity threshold).
3. The reranker specification, including reranking criteria and the mechanism by which candidate documents are scored and filtered.
4. The generator model specification, including context window size, the mechanism by which retrieved context is presented to the model, and the instruction framework governing output generation.
5. The data flow architecture: how the query is transformed into an embedding vector, how retrieval results are assembled into context, and how context is integrated with the generation prompt.
6. Boundary B1 characterisation: the semantic alignment methodology between the embedding model's vector space and the retrieval corpus index, including any calibration or alignment procedures.
7. Boundary B2 characterisation: the context fidelity methodology between reranked documents and the generator's input, including context truncation rules, ordering logic, and citation mechanisms.

Article 12 — Logging Architecture. The logging system for a RAG pipeline **SHALL** record, for each system invocation:

1. The input query as received.
2. The embedding vector produced by the embedding model.
3. The retrieval results: document identifiers, similarity scores, and retrieval rank for all candidate documents returned by the vector database.
4. The reranking results: reranked scores and the final context set presented to the generator, including any documents filtered or truncated.
5. The generator input: the complete prompt including system instructions, retrieved context, and user query as assembled for the generator.
6. The generator output: the complete response produced.

7. A unique correlation identifier linking all log entries for a single invocation across all pipeline stages.

This logging granularity enables the causal reconstruction required by Article 12(2): given a specific output, the logs must permit identification of which retrieved documents contributed to the output and whether the retrieval, reranking, or generation stage introduced the relevant content.

Article 14 — Human Oversight Architecture. Human oversight of a RAG pipeline **SHALL** be structured around the following intervention points:

1. **Pre-retrieval review:** for high-stakes decisions, the oversight architecture **SHOULD** support query review before retrieval is executed, enabling the human overseer to assess whether the query is well-formed for the retrieval corpus.
2. **Post-retrieval, pre-generation review:** the oversight architecture **SHALL** support inspection of retrieved and reranked context before it is presented to the generator. This is the primary intervention point where the human overseer can identify irrelevant, outdated, or potentially poisoned context.
3. **Post-generation review:** the oversight architecture **SHALL** support review of the generator's output before it is delivered to the end user or acted upon, with the ability to override, modify, or reject the output (Article 14(4)(d)).

The level of human oversight at each intervention point **SHALL** be proportionate to the risk level of the specific use case. For Annex III credit decisions, post-retrieval and post-generation review **SHOULD** be mandatory for decisions that produce adverse outcomes.

Article 15 — Accuracy and Robustness Specification. The declared accuracy metrics for a RAG pipeline **SHALL** include:

1. **Retrieval precision at k:** the proportion of the top-k retrieved documents that are relevant to the query, measured across a representative evaluation set.
2. **Context fidelity:** the proportion of the generated output that is faithfully grounded in the retrieved context, as opposed to hallucinated or extrapolated content.
3. **End-to-end accuracy:** the proportion of system outputs that are factually correct and decision-appropriate, measured at the system level across the intended use case.
4. **Compositional error rate:** the proportion of system errors attributable to compositional interaction (cascade failure, semantic drift) as opposed to individual component failure, measured through causal attribution analysis of error cases.

The specification **SHALL** explicitly state that end-to-end accuracy cannot be derived from the product of component-level metrics. Retrieval precision $\times$ generation accuracy $\neq$ pipeline accuracy. The declared metrics **SHALL** be derived from system-level evaluation on representative test sets that exercise the full pipeline, not from component benchmarks.

EU Compliance Series — Forthcoming

Specific evaluation set construction requirements, minimum test set sizes, statistical confidence thresholds for declared accuracy metrics, and adversarial robustness testing protocols for RAG pipelines will be provided in the CRSA-1 Financial Services Edition addressing DORA and AI Act dual-compliance requirements.

4.3 Profile 2: Agentic System in Safety-Critical Decision-Making

4.3.1 Architecture Reference

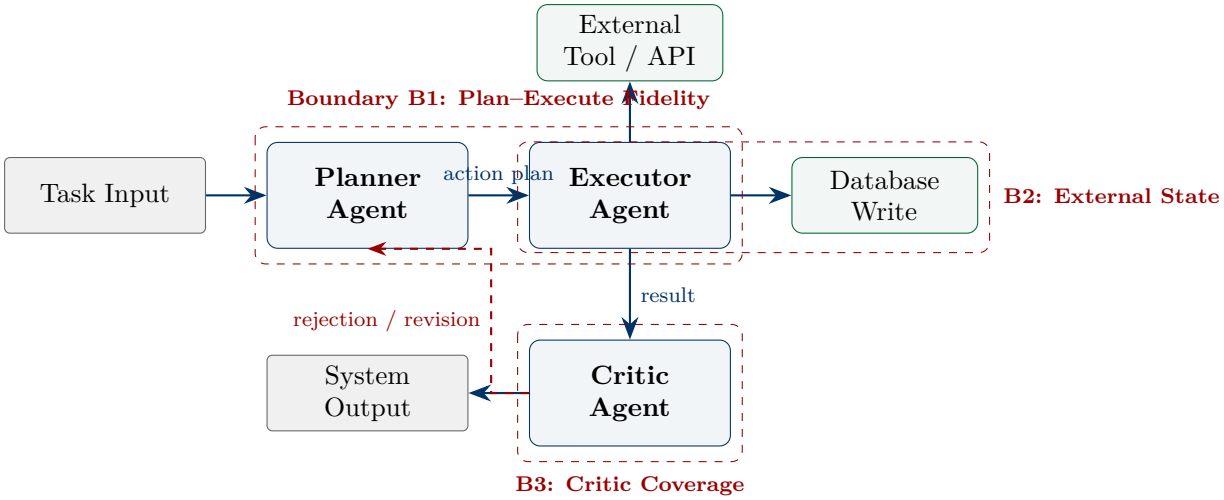

Figure 3: Agentic system architecture reference. Three compliance-critical boundaries identified: B1 (fidelity between planner intent and executor action), B2 (external state modification by executor), B3 (critic coverage of executor outputs). Dashed red line indicates the feedback loop from critic to planner.

4.3.2 Applicable Risk Categories

Risk Category	Severity	Agentic-Specific Manifestation
CRT-7 State Corruption	Critical	Executor writes to external database; interruption or failure leaves partial state. Subsequent operations inherit corrupted records.
CRT-1 Cascade Failure	High	Planner hallucinates an action step; executor faithfully executes it; critic fails to catch it because the hallucinated step is internally coherent.
CRT-5 Capability Drift	High	Dynamic tool selection expands the executor's operational scope beyond the assessed capability envelope. The system acquires capabilities not covered by the conformity assessment.
CRT-6 Context Poisoning	High	Tool output contains adversarial instructions that the executor processes as trusted context, executing unintended actions.
CRT-8 Comp. Opacity	High	The feedback loop between critic and planner creates iterative decision pathways that resist post-hoc explanation. The final output reflects multiple revision cycles whose causal contributions are non-decomposable.
CRT-4 Routing Discrim.	Medium	Planner's tool selection varies by input profile, creating differential capability pathways for different user demographics.

4.3.3 Compliance Requirements by Article

Article 9 — Risk Management. The risk management system for an agentic architecture **SHALL** identify and assess the following compositional risk pathways:

1. **Plan fidelity failure:** the risk that the executor interprets the planner's action specification incorrectly or incompletely, producing actions that diverge from the planner's intent. This is a cascade failure (CRT-1) originating at Boundary B1.
2. **Critic coverage gaps:** the risk that the critic agent fails to evaluate executor outputs against all relevant safety criteria, permitting unsafe actions to reach the system output.

The risk management system **SHALL** characterise the critic's coverage envelope: the set of failure modes the critic is designed to detect, and the residual failure modes outside its coverage.

3. **Feedback loop instability:** the risk that the critic-to-planner feedback loop produces oscillating or divergent revision cycles, where successive plan revisions do not converge on a stable action sequence. The risk management system **SHALL** define maximum iteration bounds for the feedback loop.
4. **External state modification:** the risk that the executor modifies external system state (databases, APIs, records) in ways that cannot be reversed upon system interruption or error detection (CRT-7 at Boundary B2).
5. **Dynamic capability expansion:** the risk that the planner selects tools or capabilities not included in the conformity assessment, expanding the system's operational envelope beyond its assessed boundary (CRT-5).

Risk management for agentic systems is fundamentally dynamic: the risk profile changes with each action-observation cycle. The risk management system **SHALL** operate continuously during inference, not merely at design time. This imposes requirements on runtime monitoring infrastructure that static risk assessment methodologies do not address.

Article 11 / Annex IV 2(c) — Architecture Documentation. The technical documentation for an agentic system **SHALL** include, at minimum:

1. The planner agent specification, including the planning methodology (chain-of-thought, tree-of-thought, ReAct, or other), the action space definition (the set of actions available to the planner), and the planning horizon (number of steps the planner considers).
2. The executor agent specification, including the execution mechanism, the tool-calling interface, and the set of external systems the executor can interact with.
3. The critic agent specification, including the evaluation criteria, the coverage envelope, and the rejection/revision mechanism.
4. The feedback loop specification: maximum iteration count, convergence criteria, and termination conditions.
5. The tool inventory: a complete enumeration of tools, APIs, and external systems accessible to the executor, including read/write permissions and state modification capabilities for each.
6. Boundary B1 characterisation: how the planner's output is translated into executor instructions, including any parsing, validation, or constraint enforcement at this boundary.
7. Boundary B2 characterisation: how the executor interacts with external systems, including transactional guarantees, rollback capabilities, and state consistency mechanisms.
8. Boundary B3 characterisation: the critic's coverage envelope, including the safety criteria evaluated, the evaluation methodology, and the conditions under which the critic can override the executor.
9. Annex IV 2(f) predetermined changes: a specification of foreseeable tool additions, model version updates, and capability extensions, with the parametric boundaries within which each change remains non-substantial.

Article 12 — Logging Architecture. The logging system for an agentic architecture **SHALL** record, for each system invocation:

1. The task input as received.
2. Each planner iteration: the action plan produced, including the reasoning chain and selected tools.
3. Each executor action: the tool called, the parameters passed, the external system response received, and any state modifications executed.

4. Each critic evaluation: the criteria assessed, the verdict (accept/reject/revise), and the reasoning supporting the verdict.
5. Each feedback loop iteration: the revision requested by the critic, the planner's revised plan, and the iteration count.
6. The final system output as delivered.
7. A unique correlation identifier linking all log entries across all agents and iterations for a single task invocation.
8. For each external state modification: a before-state snapshot, the modification applied, and the after-state, enabling rollback reconstruction.

The logging granularity for agentic systems exceeds that for any other architecture type. The feedback loop between critic and planner means that the causal pathway from input to output is non-linear: the output reflects multiple planning-execution-evaluation cycles. Logs **MUST** capture the full iteration history to satisfy Article 12(2)'s traceability requirement.

Article 14 — Human Oversight Architecture. Human oversight of an agentic system **SHALL** address the unique challenge that agents operate autonomously at machine speed while human oversight operates at human speed. The oversight architecture **SHALL** implement:

1. **Pre-execution approval gates:** for safety-critical actions (external state modifications, irreversible decisions), the architecture **SHALL** require explicit human approval before the executor proceeds. The set of actions requiring pre-execution approval **SHALL** be defined in the risk management plan and documented under Annex IV.
2. **Asynchronous monitoring dashboard:** a real-time visibility layer providing the human overseer with the current state of the planning-execution-evaluation cycle, including the active plan, pending actions, completed actions, and critic evaluations.
3. **Safe halt mechanism:** the "stop button" required by Article 14(4)(e) **SHALL** implement the following sequence: (i) immediately prevent the planner from issuing new action steps; (ii) allow the currently executing action to complete if it is non-state-modifying, or abort and roll back if it is state-modifying; (iii) place the system in a defined safe state with all external system interactions cleanly terminated.
4. **Rollback capability:** upon halt, the oversight architecture **SHALL** support reversal of state modifications executed during the current task invocation, to the extent that external systems support transactional rollback (Article 14(4)(d)).

The safe halt mechanism directly addresses the CRT-7 (State Corruption) risk that is rated Critical for agentic architectures. Without transactional composition guarantees, the Article 14(4)(e) "stop button" requirement cannot be satisfied for agentic systems that modify external state.

Article 15 — Accuracy and Robustness Specification. The declared accuracy metrics for an agentic system **SHALL** include:

1. **Task completion rate:** the proportion of task inputs for which the system produces a correct and complete output, measured across a representative evaluation set exercising the full planning-execution-evaluation cycle.
2. **Plan fidelity:** the proportion of executor actions that faithfully implement the planner's intent, measured through automated comparison of planned actions and executed actions.
3. **Critic detection rate:** the proportion of executor errors that the critic successfully identifies, measured against a test set of known error cases injected into the execution pipeline.
4. **Feedback convergence rate:** the proportion of task invocations in which the critic-planner feedback loop converges within the defined iteration bound, versus those terminated by the iteration limit without convergence.
5. **Compositional error rate:** the proportion of task failures attributable to inter-agent interaction (cascade failure between planner and executor, critic coverage gaps, feedback

instability) as opposed to individual component failure.

Accuracy for an agentic system is context-dependent: the system may take different valid action sequences for identical inputs. The declared metrics **SHALL** specify the evaluation methodology for determining output correctness when multiple valid action pathways exist, including the acceptance criteria for "correct and complete" task completion.

Robustness testing **SHALL** include adversarial scenarios targeting each compliance-critical boundary:

1. **B1 adversarial:** inputs designed to cause the planner to generate action plans that the executor will interpret in unintended ways.
2. **B2 adversarial:** tool outputs containing adversarial content designed to manipulate the executor's subsequent actions (CRT-6 testing).
3. **B3 adversarial:** executor outputs designed to evade the critic's evaluation criteria while containing safety-relevant errors.

EU Compliance Series — Forthcoming

Specific evaluation set construction requirements for agentic systems, including task complexity taxonomies, adversarial scenario libraries, and minimum test coverage requirements for critic evaluation, will be provided in the CRSA-1 Medical Devices Edition addressing MDR and AI Act dual-conformity requirements for agentic diagnostic systems.

4.4 Profile 3: Ensemble System in Automated Assessment

4.4.1 Architecture Reference

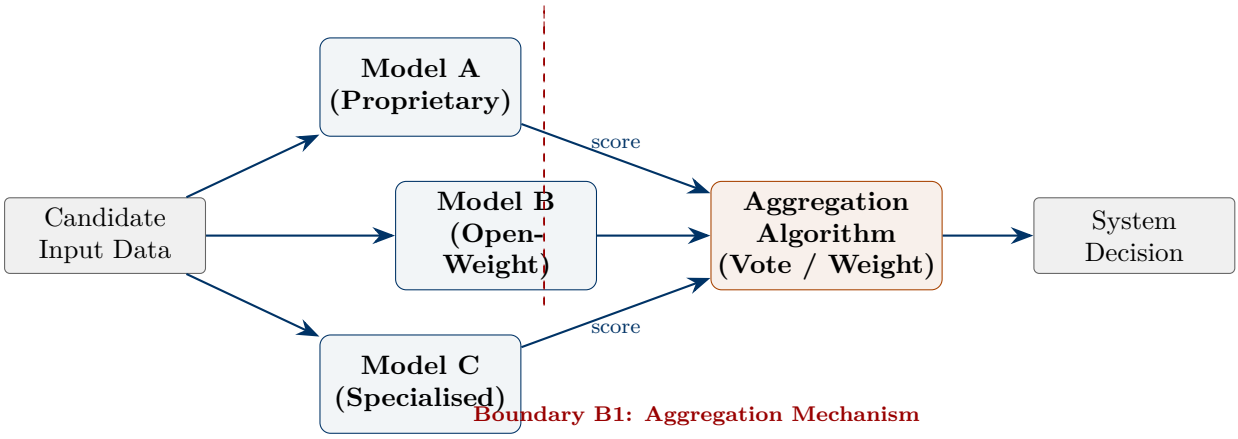

Figure 4: Ensemble system architecture reference. One compliance-critical boundary identified: B1 (the aggregation mechanism through which individual model outputs are combined into the system decision). Components may originate from different providers with different training data, bias profiles, and performance characteristics.

4.4.2 Applicable Risk Categories

Risk Category	Severity	Ensemble-Specific Manifestation
CRT-3 Aggregation Bias	High	Model A biased against demographic X, Model C biased against demographic Y. Voting aggregation may produce compound discrimination against both groups, neither group, or one group depending on thresholds and weighting—a non-linear bias interaction invisible to single-model fairness testing.
CRT-8 Comp. Opacity	High	A non-linear weighted vote among three models produces a decision whose causal attribution to individual model contributions is mathematically non-trivial. Standard feature attribution methods apply within models but not across the aggregation boundary.
CRT-2 Semantic Drift	Medium	Component models trained on different data distributions interpret the same input differently. The aggregation mechanism combines outputs whose underlying semantic interpretations diverge.
CRT-1 Cascade Failure	Low–Medium	One model produces a confident but incorrect output that dominates the aggregation when weighting favours high-confidence scores, overriding correct outputs from other models.

4.4.3 Compliance Requirements by Article

Article 9 — Risk Management. The risk management system for an ensemble architecture **SHALL** identify and assess the following compositional risk pathways:

1. **Aggregation bias emergence:** the risk that individually fair models produce discriminatory aggregate decisions through the interaction of their respective bias profiles with the aggregation mechanism (CRT-3). The risk assessment **SHALL** include disaggregated fairness analysis at both the component level and the system level, with explicit evaluation of whether the aggregation mechanism preserves, amplifies, or introduces bias.
2. **Confidence dominance:** the risk that one model's high-confidence output disproportionately influences the aggregate decision, effectively reducing the ensemble to a single-model system under certain input conditions.
3. **Correlated failure:** the risk that component models trained on overlapping or related datasets exhibit correlated errors, reducing the diversity benefit that motivates ensemble design. When all models fail on the same input, the aggregation mechanism amplifies rather than corrects the error.
4. **Component provenance asymmetry:** the risk arising when component models originate from different providers with different update cycles, documentation quality, and cooperation levels under Article 25(4). An upstream provider's silent model update may alter the ensemble's balance without the system provider's knowledge.

Article 10 — Data Governance. Data governance for ensemble systems **SHALL** address the compound data lineage challenge:

1. Each component model's training, validation, and testing data **SHALL** individually satisfy Article 10(2) requirements.
2. The system provider **SHALL** document the data governance status of each component, including whether the provider has obtained sufficient information from upstream suppliers (per Article 25(4)) to verify data compliance.

3. The bias examination required by Article 10(2)(f) **SHALL** extend to the aggregation mechanism: the provider **SHALL** assess how the interaction of component-level training data biases manifests through the specific aggregation algorithm employed.
4. Where an open-weight component was trained on datasets outside the system provider's control, the provider **SHALL** document the known data governance limitations and the compensating measures applied (such as system-level bias testing that captures aggregation effects).

Article 11 / Annex IV 2(c) — Architecture Documentation. The technical documentation for an ensemble system **SHALL** include, at minimum:

1. The specification of each component model, including provider, version, training domain, and known performance characteristics.
2. The aggregation algorithm specification: voting mechanism (majority vote, weighted vote, soft vote, stacking), weighting scheme (fixed, adaptive, input-dependent), tie-breaking rules, and confidence threshold parameters.
3. The rationale for ensemble composition: why these specific models were selected, what diversity benefit each provides, and how the composition is expected to outperform individual components.
4. Boundary B1 characterisation: how individual model outputs are normalised, scaled, or transformed before aggregation, and how the aggregation mechanism produces the final decision.
5. The Article 25(4) agreement status for each third-party component, including the scope of information obtained from each upstream provider.

Article 12 — Logging Architecture. The logging system for an ensemble architecture **SHALL** record, for each system invocation:

1. The input as presented to all component models.
2. Each component model's individual output, including raw scores, confidence values, and classification labels.
3. The aggregation computation: the weights applied, the normalisation performed, and the aggregation formula executed.
4. The final system decision and the margin by which it was reached (for voting systems, the vote count; for weighted systems, the weighted score differential).
5. A unique correlation identifier linking all component outputs and the aggregation computation for a single invocation.

The question of whether logs **MUST** record individual model outputs or may record only the aggregate decision has no regulatory answer in the current framework. This specification resolves the ambiguity: individual model outputs **SHALL** be logged. Article 12(2) requires traceability sufficient to identify risk situations and facilitate post-market monitoring. Risk situations in ensemble systems (aggregation bias, confidence dominance, correlated failure) can only be identified through analysis of individual component outputs relative to the aggregate decision.

Article 13 — Transparency. The instructions of use for an ensemble system **SHALL** disclose to deployers:

1. That the system employs an ensemble architecture combining multiple models.
2. The number of component models and their general characteristics (without requiring disclosure of proprietary model internals).
3. The aggregation mechanism in sufficient detail to enable the deployer to understand how individual model outputs are combined into the system decision.

4. The known limitations of the aggregation mechanism, including conditions under which confidence dominance, correlated failure, or aggregation bias may occur.

This disclosure is necessary to satisfy Article 13(1)'s requirement that deployers can "interpret the system's output and use it appropriately." A deployer who believes they are using a single-model system cannot interpret ensemble outputs appropriately.

Article 14 — Human Oversight Architecture. Human oversight of an ensemble system **SHALL** implement:

1. **Disagreement alerting:** when component models produce substantially divergent outputs (high disagreement among voters), the oversight architecture **SHALL** flag the invocation for human review. High disagreement indicates an input where the ensemble's diversity benefit is most needed but also where compositional uncertainty is highest.
2. **Margin-based escalation:** when the aggregate decision is reached by a narrow margin (close vote, small weighted score differential), the oversight architecture **SHOULD** escalate to human review, particularly for Annex III decisions affecting fundamental rights.
3. **Override capability:** the human overseer **SHALL** be able to override the aggregate decision, with the override and its reasoning recorded in the Article 12 logging system.

Article 15 — Accuracy and Robustness Specification. The declared accuracy metrics for an ensemble system **SHALL** include:

1. **System-level accuracy:** measured across the representative evaluation set, reflecting the aggregated output.
2. **Component-level accuracy:** each model's individual accuracy, providing the baseline against which the ensemble's compositional accuracy gain is measured.
3. **Diversity index:** a measure of the independence of component model errors, quantifying the degree to which the ensemble benefits from model diversity (e.g., Q-statistic, disagreement measure, or correlation coefficient of errors).
4. **Aggregation robustness:** the system's accuracy under adversarial conditions targeting the aggregation mechanism, including inputs designed to exploit confidence dominance or trigger correlated failure across components.

4.5 Profile 4: Cascade/Routing System in Public Services

4.5.1 Architecture Reference

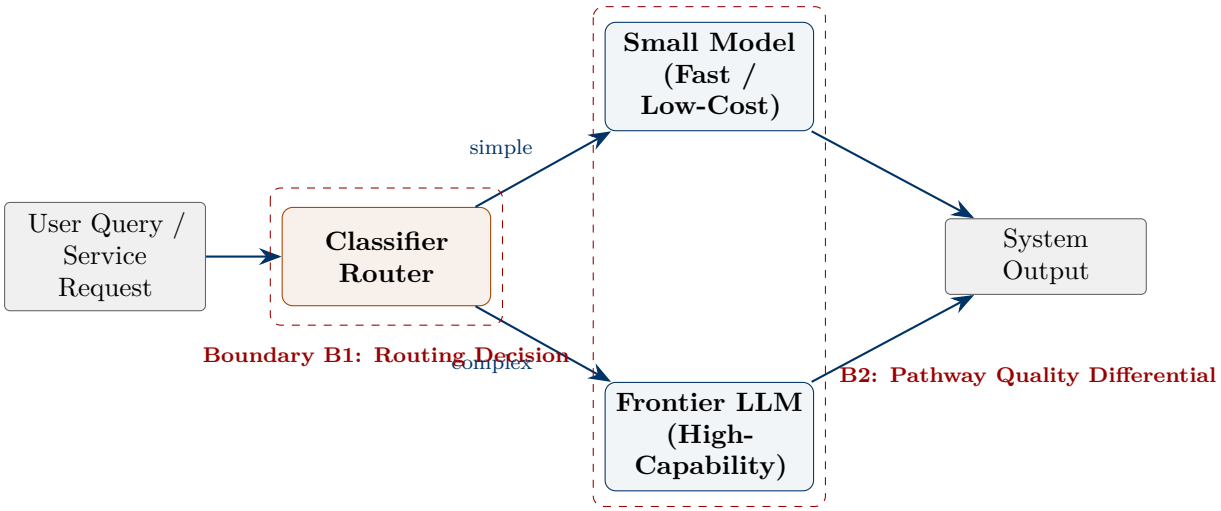

Figure 5: Cascade/routing system architecture reference. Two compliance-critical boundaries identified:

B1 (the routing decision determining which model processes each input) and B2 (the quality differential between processing pathways).

4.5.2 Applicable Risk Categories

Risk Category	Severity	Cascade-Specific Manifestation
CRT-4 Routing Discrim.	High	Router misclassifies speakers of certain dialects or demographic profiles, systematically directing them to the lower-capability pathway. The routing architecture creates structural discrimination.
CRT-1 Cascade Failure	High	Router misclassifies query complexity, sending a complex query to the small model. The small model produces a plausible but incorrect response that the system delivers with the same confidence framing as frontier model outputs.
CRT-8 Comp. Opacity	Medium	Users and deployers cannot determine which model processed their request, making output interpretation and error attribution impossible without compositional logging.
CRT-5 Capability Drift	Medium	Router thresholds calibrated for one model version become miscalibrated when the downstream model is updated, routing queries to the wrong pathway.

4.5.3 Compliance Requirements by Article

Article 9 — Risk Management. The risk management system for a cascade/routing architecture **SHALL** identify and assess the following compositional risk pathways:

1. **Misrouting:** the risk that the classifier routes queries to the wrong processing pathway, resulting in either under-service (complex query to small model) or unnecessary cost (simple query to frontier model). The risk assessment **SHALL** quantify the misrouting rate disaggregated by input characteristics.
2. **Routing discrimination:** the risk that the routing decision correlates with protected characteristics, creating differential service quality for demographic groups (CRT-4). The risk assessment **SHALL** evaluate routing fairness jointly with pathway quality differential—not merely whether the router is fair in isolation.
3. **Threshold sensitivity:** the risk that small changes in the routing threshold produce large changes in system behaviour, creating fragile system performance around the decision boundary.
4. **Pathway calibration drift:** the risk that the routing thresholds, calibrated for specific model versions, become misaligned when downstream models are updated (CRT-5). The risk management system **SHALL** define recalibration triggers linked to downstream model updates.

Article 11 / Annex IV 2(c) — Architecture Documentation. The technical documentation for a cascade/routing system **SHALL** include, at minimum:

1. The classifier/router specification, including the classification methodology, feature set, and training data.
2. The routing threshold specification: the decision boundary, the confidence threshold, and the tie-breaking rules.
3. The specification of each downstream model, including capability characterisation, intended query profile, and performance envelope.
4. The quality differential characterisation: a quantified comparison of the accuracy, capability, and limitation profile of each processing pathway.

5. Boundary B1 characterisation: how the routing decision is made, what input features influence the decision, and how the router's confidence is calibrated.
6. Boundary B2 characterisation: the performance differential between pathways, including scenarios where the differential is most pronounced.

Article 12 — Logging Architecture. The logging system for a cascade/routing architecture **SHALL** record, for each system invocation:

1. The input query as received.
2. The router's classification decision, including the confidence score and the routing pathway selected.
3. The downstream model that processed the query (model identifier and version).
4. The downstream model's output.
5. A unique correlation identifier linking the routing decision to the downstream processing and output.

Logging of the routing decision and pathway selection is essential for detecting CRT-4 (Routing Discrimination) through post-market monitoring. Without this data, disaggregated analysis of service quality across demographic groups is impossible.

Article 15 — Accuracy and Robustness Specification. The declared accuracy metrics for a cascade/routing system **SHALL** address the three-dimensional specification problem:

1. **Routing accuracy:** the proportion of queries correctly classified by the router as simple or complex, measured against a labelled evaluation set.
2. **Pathway-specific accuracy:** each downstream model's accuracy on its intended query profile—the small model's accuracy on queries classified as simple, and the frontier model's accuracy on queries classified as complex.
3. **End-to-end system accuracy:** the combined accuracy reflecting routing accuracy × pathway-specific accuracy, weighted by the routing distribution. This metric **SHALL** be declared as the system's accuracy under Article 15(2).
4. **Misrouting impact:** the accuracy degradation when queries are misrouted—specifically, the small model's accuracy on queries that should have been routed to the frontier model. This quantifies the cost of routing error and informs the risk management assessment under Article 9.

The specification **SHALL** explicitly state that system-level accuracy is not uniform across inputs: users whose queries are routed to the small model receive a different accuracy level than users whose queries are routed to the frontier model. This population-dependent accuracy **SHALL** be disclosed in the instructions of use under Article 13(3)(b).

Article 27 — Fundamental Rights Impact Assessment. For cascade/routing systems deployed by public bodies or in Annex III use cases, the fundamental rights impact assessment **SHALL** specifically evaluate:

1. Whether the routing decision disproportionately directs members of protected groups to the lower-capability pathway.
2. Whether the quality differential between pathways creates materially different outcomes for affected persons based on routing, and whether those outcome differentials correlate with protected characteristics.
3. Whether the routing architecture constitutes indirect discrimination under the Equal Treatment Directives (2000/43/EC, 2000/78/EC) by creating an apparently neutral mechanism that disproportionately disadvantages protected groups.

4.6 Profile 5: MCP-Based Multi-Agent System in Enterprise Operations

4.6.1 Architecture Reference

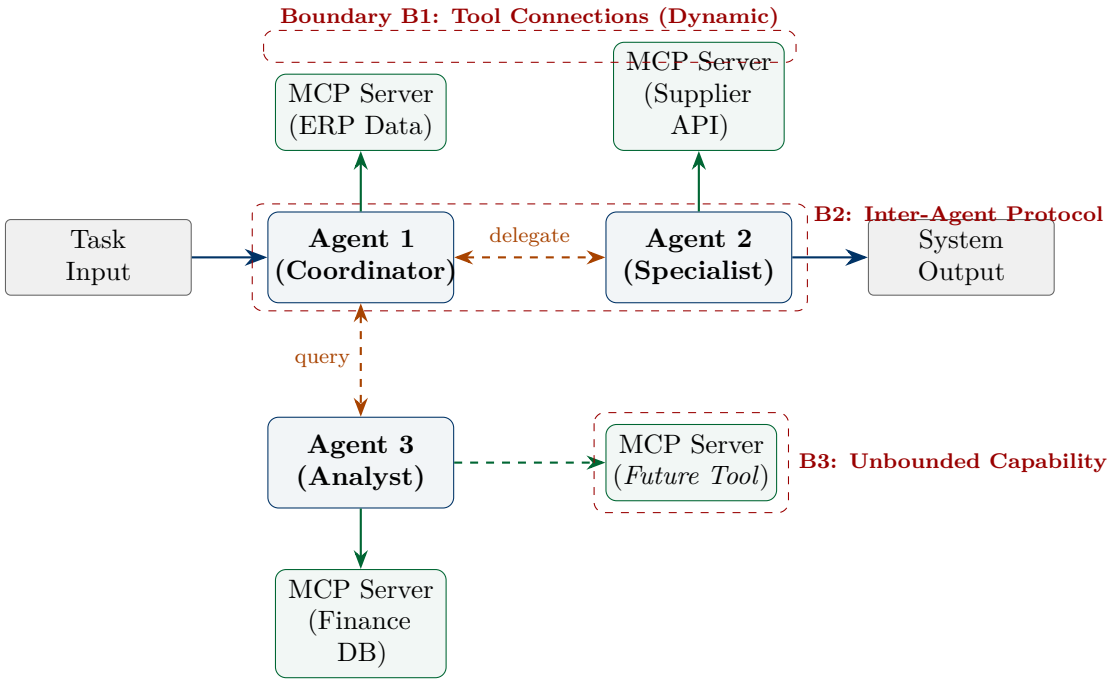

Figure 6: MCP-based multi-agent system architecture reference. Three compliance-critical boundaries identified: B1 (dynamic tool connections via MCP servers), B2 (inter-agent communication protocol), B3 (unbounded capability space from future tool connections, shown dashed). The dashed MCP connection to "Future Tool" represents capability drift risk.

4.6.2 Applicable Risk Categories

Risk Category	Severity	MCP-Specific Manifestation
CRT-5 Capability Drift	Critical	Connecting a new MCP server post-deployment grants the system access to data sources and actions not covered by the conformity assessment. The system that was assessed is not the system that operates.
CRT-6 Context Poisoning	Critical	Any connected MCP data source is a potential injection vector. Adversarial content in a supplier catalog, ERP field, or external database is retrieved by an agent and processed as trusted context.
CRT-7 State Corruption	High	Multiple agents writing to multiple external systems via MCP create distributed state management complexity. Interruption leaves partial state across multiple external systems.
CRT-1 Cascade Failure	High	Agent 1 delegates a subtask to Agent 2 based on incorrect information from an MCP data source. Agent 2 executes faithfully, propagating the error through to the system output.
CRT-8 Comp. Opacity	High	Multi-agent delegation chains with dynamic tool access produce decision pathways that are architecturally non-deterministic: the same input may produce different agent interaction patterns depending on MCP data state.
CRT-4 Routing Discrim.	Medium	Agent 1's delegation decisions may route different types of requests to different specialist agents with different capability profiles, creating differential service quality.

4.6.3 Compliance Requirements by Article

Article 9 — Risk Management. The risk management system for an MCP-based multi-agent architecture **SHALL** identify and assess the following compositional risk pathways:

1. **Unbounded capability expansion:** the risk that post- deployment MCP tool connections grant the system capabilities not assessed in the conformity assessment (CRT-5). The risk management system **SHALL** maintain a runtime capability registry and evaluate every new tool connection against the conformity assessment scope before activation.
2. **Multi-vector context poisoning:** the risk that adversarial content in any connected data source is retrieved and acted upon (CRT-6). The attack surface scales linearly with the number of connected MCP servers, making multi-agent systems disproportionately exposed relative to single-model deployments.
3. **Distributed state corruption:** the risk that concurrent agent actions modify multiple external systems, and that interruption or failure leaves distributed state inconsistent (CRT-7). The risk is compounded when agents operate asynchronously on shared external resources.
4. **Inter-agent delegation fidelity:** the risk that the coordinating agent's task delegation is misinterpreted by specialist agents, producing actions that diverge from the coordinator's intent. This is a multi-agent variant of CRT-1 (Cascade Failure) operating across the inter-agent protocol (Boundary B2).
5. **Emergent autonomy:** the risk that the combination of multiple agents with multiple tool connections produces a system whose aggregate autonomous capability exceeds any individual agent's autonomy level, potentially implicating the systemic risk considerations under Article 51 and Annex XIII.

Article 11 / Annex IV — Architecture Documentation. The technical documentation for an MCP-based multi-agent system **SHALL** include, at minimum:

1. The specification of each agent, including its role, capability envelope, and the scope of actions it can perform.
2. The inter-agent communication protocol: how agents delegate tasks, share information, and coordinate actions, including message formats, routing logic, and conflict resolution mechanisms.
3. The MCP server inventory: a complete enumeration of connected MCP servers at the time of conformity assessment, including the data sources and capabilities each server exposes, and the read/write permissions granted.
4. The capability governance policy: the criteria and procedures by which new MCP server connections are evaluated, approved, and documented, including the compliance review required before activation.
5. Boundary B1 characterisation: the MCP connection interface, including authentication, authorisation, input validation, and output sanitisation for each connected server.
6. Boundary B2 characterisation: the inter-agent protocol, including message validation, delegation scope constraints, and agent-to-agent trust boundaries.
7. Boundary B3 specification: the pre-documented scope of foreseeable tool additions under Annex IV 2(f), including the categories of MCP servers that may be connected without triggering re-assessment, and the categories that constitute a substantial modification requiring re-assessment.
8. Annex IV 1(b) — system boundary declaration: a clear determination of which MCP-connected tools and data sources fall within the AI system boundary (and are therefore covered by the conformity assessment) and which are "other systems" at the boundary (covered by the interaction documentation but not by the system-level assessment).

Article 12 — Logging Architecture. The logging system for an MCP-based multi-agent architecture **SHALL** record, for each system invocation:

1. The task input as received by the coordinating agent.
2. Each inter-agent delegation: the delegating agent, the receiving agent, the task specification, and the delegation rationale.
3. Each MCP server interaction: the agent initiating the interaction, the MCP server accessed, the query or action executed, the data returned, and any state modifications applied.
4. Each agent's reasoning or processing output.
5. The final system output and the agent that produced it.
6. A distributed correlation identifier linking all agent actions, MCP interactions, and inter-agent communications for a single task invocation.
7. The MCP server inventory at the time of invocation, enabling post-hoc verification that the operational system matched the assessed system.

Logging the MCP server inventory per invocation is specific to multi-agent systems and directly addresses the CRT-5 (Capability Drift) detection requirement: comparing the operational inventory against the assessed inventory enables continuous conformity verification.

Article 14 — Human Oversight Architecture. Human oversight of an MCP-based multi-agent system **SHALL** implement:

1. **Delegation approval gates:** for safety-critical delegations (actions involving external state modification, financial transactions, or decisions affecting fundamental rights), the oversight architecture **SHALL** require human approval before the delegation is executed.
2. **MCP connection approval:** new MCP server connections **SHALL** require human authorisation, with the human verifying that the connection falls within the pre-documented scope of Annex IV 2(f) predetermined changes or initiating a compliance review if it does not.
3. **Distributed safe halt:** the "stop button" **SHALL** propagate halt signals to all active agents simultaneously, prevent new MCP interactions from initiating, allow in-progress non-state-modifying MCP interactions to complete, abort and roll back in-progress state-modifying MCP interactions, and produce a complete halt state log capturing the state of all agents and MCP connections at the point of interruption.
4. **Scope monitoring:** continuous verification that the system's operational capability envelope (the set of active MCP connections and agent capabilities) matches the assessed capability envelope, with automatic alerting when divergence is detected.

Article 15 — Accuracy and Robustness Specification. The declared accuracy metrics for an MCP-based multi-agent system **SHALL** include:

1. **Task completion rate:** as defined in Profile 2 (Section 4.3).
2. **Delegation fidelity:** the proportion of inter-agent delegations where the receiving agent's action faithfully implements the coordinating agent's specification.
3. **MCP data accuracy:** the proportion of MCP data retrievals that return accurate, current, and uncompromised data, measured through periodic validation against known-good reference data.
4. **Compositional error rate:** task failures attributable to inter-agent or agent-MCP interaction failures as opposed to individual agent or individual MCP server failures.

Robustness testing **SHALL** include:

1. **Multi-vector context poisoning:** adversarial content injected into each connected MCP data source to test the system's resilience to indirect prompt injection through data retrieval.
2. **Inter-agent protocol manipulation:** adversarial delegation messages designed to cause specialist agents to take actions outside their intended scope.

3. **MCP server unavailability:** graceful degradation testing when one or more MCP servers become unavailable mid-operation, verifying that the system halts safely rather than producing incomplete or corrupted outputs.

Article 43(4) — Substantial Modification Assessment. For MCP-based systems, the following **SHALL** be assessed as potential substantial modifications requiring re-assessment:

1. Connecting an MCP server that exposes data sources or capabilities outside the categories pre-documented under Annex IV 2(f).
2. Adding a new agent to the multi-agent system.
3. Modifying the inter-agent communication protocol.
4. Granting an existing agent write access to an external system previously accessible in read-only mode.

Connecting an MCP server within a pre-documented category (e.g., adding a second ERP data source when ERP data sources were pre-documented) **SHOULD** not require re-assessment, provided the system provider verifies that the new connection falls within the assessed parametric boundaries.

EU Compliance Series — Forthcoming

MCP-specific conformity assessment protocols, including capability governance audit procedures, distributed halt testing requirements, and multi-vector adversarial testing specifications, will be provided in the CRSA-1 Critical Infrastructure Edition addressing NIS2 and AI Act dual-compliance requirements for multi-agent operational systems.

5 Conformity Assessment Methodology for Composed Systems

5.1 Purpose and Regulatory Basis

Article 43 of Regulation (EU) 2024/1689 requires high-risk AI systems to undergo conformity assessment before placement on the European market. The assessment may follow the internal control procedure (Annex VI) or the third-party notified body procedure (Annex VII), depending on the system's classification. In both cases, the assessment must demonstrate compliance with the full set of Chapter III, Section 2 requirements.

For composed AI systems, the conformity assessment faces a structural obstacle identified in Section 2 and flagged as CRITICAL: no harmonized standard or notified body methodology exists for assessing systems composed of multiple AI components from different providers. The AI Act lacks a formalised mechanism for modular certification. Component-level conformity evidence cannot be aggregated to demonstrate system-level compliance.

This section provides the conformity assessment methodology that closes the gap. It is structured to serve both system providers performing internal control assessment (Annex VI) and notified bodies conducting third-party assessment (Annex VII), providing a common evidence framework applicable across both procedures.

5.2 Assessment Scope Definition

5.2.1 System Boundary Determination

Before any conformity assessment can proceed, the provider **MUST** define the system boundary: the set of components, data sources, tools, and interfaces that constitute the AI system being assessed. For composed systems, this determination is architecturally significant and legally consequential.

The system boundary **SHALL** be defined using the following criteria:

1. **Processing path inclusion:** any component through which data flows during the system's inference process—from input reception to output delivery—is within the system boundary. This includes retrieval models, rerankers, generators, routers, planners, executors, critics, and aggregation algorithms.

2. **State modification inclusion:** any external system that the composed AI system can modify during operation (databases, APIs, records) is within the system boundary to the extent of the modification capability. Read-only external data sources are at the boundary; write-capable external systems are within it.

3. **Tool and MCP inclusion:** for MCP-based systems, all MCP servers connected at the time of assessment are within the system boundary. MCP servers connected after assessment are outside the boundary unless pre-documented under Annex IV 2(f) as predetermined changes within specified categories.

4. **Orchestration inclusion:** the orchestration logic (routing algorithms, delegation protocols, feedback loops, voting mechanisms) is within the system boundary regardless of whether it is implemented as a separate software component or embedded within a model's prompt framework.

The system boundary determination **SHALL** be documented in the Annex IV technical documentation as part of the general system description required by Annex IV 1(b). Annex IV 1(b) requires documentation of how the system "interacts with, or can be used to interact with, hardware or software, including with other AI systems, that are not part of the AI system itself." The system boundary determination defines the scope of "the AI system itself" and, by exclusion, identifies the external systems at the boundary.

5.2.2 Provider Determination

Under Articles 3(3) and 25, the provider of the composed system is the entity that assembles the components and places the resulting system on the market or puts it into service under its own name or trademark. The conformity assessment methodology requires explicit provider determination as a prerequisite:

1. **Single-provider assembly:** where one entity develops or assembles the composed system from its own and third-party components, that entity is the provider and bears all Article 16 obligations.

2. **Deployer-turned-provider:** where a deployer integrates third-party components into a pipeline, applies custom orchestration, or modifies the intended purpose of component models to create a high-risk system, the deployer becomes the provider under Article 25(1). The assessment must be performed by the new provider.

3. **Substantial modification reclassification:** where an entity makes a substantial modification to an existing composed system (component swap, tool addition, routing threshold change not pre-documented as a predetermined change), the modifying entity becomes the new provider and must perform a new conformity assessment.

In all cases, the provider identified through this determination is the entity responsible for the conformity assessment, the EU declaration of conformity (Article 47), and the registration in the EU database (Article 49).

5.3 Evidence Requirements

The conformity assessment for a composed system requires evidence across six domains, each mapped to the Chapter III, Section 2 requirements and the compositional risk taxonomy established in Section 3. The evidence domains correspond to the five critical gaps identified in the Section 2 summary plus the compositional risk management obligation.

5.3.1 Domain 1: Compositional Risk Management Evidence

Satisfies Article 9. Addresses all CRT categories.

1. **Compositional risk register:** a documented register of all compositional risks identified through application of the CRT taxonomy (Section 3) to the specific architecture. For each risk, the register **SHALL** record the CRT category, the severity assessment, the affected components, the detection methodology deployed, and the mitigation measures implemented.

2. **Interaction effects assessment:** documented analysis of the "effects and possible interaction" between Chapter III requirements as required by Article 9(4), specific to the composed architecture. This assessment **SHALL** identify where satisfying one requirement creates tension with another—for example, where logging granularity sufficient for traceability may compromise data minimisation under GDPR, or where accuracy optimisation in one component may degrade robustness in another.

3. **Residual compositional risk disclosure:** documented residual risks—compositional risks that remain after mitigation— for communication to deployers under Article 9(7). These **SHALL** include compositional residual risks that are specific to the composed architecture and absent from individual component risk profiles.

5.3.2 Domain 2: Architecture Documentation Evidence

Satisfies Article 11 and Annex IV 2(c). Addresses CRT-8 Compositional Opacity.

1. **Architecture specification:** the system architecture documentation required by Annex IV 2(c), structured according to the applicable architecture profile in Section 4. This documentation **SHALL** describe how components "build on or feed into each other and integrate into the overall processing."

2. **Third-party component documentation:** the documentation of pre-trained systems or tools provided by third parties as required by Annex IV 2(a), including the scope of information obtained from each upstream provider under Article 25(4) agreements.

3. **Boundary characterisation:** for each compliance-critical boundary identified in the architecture profile, a documented characterisation of the data transformation, validation, and integrity mechanisms operating at that boundary.

4. **Predetermined change specification:** the documentation of foreseeable component changes under Annex IV 2(f), including the parametric boundaries within which each change type remains non-substantial. This specification **SHALL** be sufficiently precise that an auditor can determine, for any given component change, whether it falls within or outside the pre-documented scope.

5.3.3 Domain 3: Logging and Traceability Evidence

Satisfies Article 12. Addresses CRT-1 and CRT-7.

1. **Logging architecture specification:** documentation of the logging system, structured according to the Article 12 requirements in the applicable architecture profile (Section 4).

2. **Cross-component correlation demonstration:** empirical evidence that the logging system can reconstruct the complete processing path for any given system output, from input through all intermediate component interactions to final output. This **SHALL** be demonstrated through sample trace reconstructions from the logging infrastructure.

3. **Risk identification capability:** evidence that the logging system can detect the compositional failure modes identified in the risk register (Domain 1), including sample detection scenarios for each applicable CRT category.

5.3.4 Domain 4: Compositional Accuracy and Robustness Evidence

Satisfies Article 15. Addresses CRT-1, CRT-2, CRT-3, and CRT-6.

1. **System-level accuracy evaluation:** empirical evaluation of the composed system's end-to-end accuracy, measured on a representative evaluation set exercising the full pipeline. The evaluation **SHALL** be conducted at the system level, not derived from component-level metrics.

2. **Component contribution analysis:** analysis demonstrating how each component contributes to system-level accuracy and where compositional interaction degrades accuracy relative to component baselines. This analysis **SHALL** identify whether system-level errors originate from individual component failure or compositional interaction (cascade failure, semantic drift).

3. **Compositional robustness testing:** adversarial testing targeting the compliance-critical boundaries identified in the architecture profile. For each boundary, the testing **SHALL** include the adversarial scenarios specified in the applicable Section 4 profile. Results **SHALL** document the system's resilience at each boundary.

4. **Compositional cybersecurity assessment:** evaluation of the system's resilience to compositional attack vectors (CRT-6 Context Poisoning), including indirect prompt injection through retrieval corpora, tool outputs, and inter-agent communication channels. The assessment **SHALL** cover each connected data source and tool within the system boundary.

5.3.5 Domain 5: Human Oversight and Safe Interruption Evidence

Satisfies Article 14. Addresses CRT-7 State Corruption.

1. **Oversight architecture specification:** documentation of the human oversight mechanisms, structured according to the Article 14 requirements in the applicable architecture profile (Section 4).

2. **Safe halt demonstration:** empirical evidence that the system's "stop button" mechanism (Article 14(4)(e)) produces a defined safe state when activated during system operation. For agentic and MCP-based systems, this **SHALL** include demonstration of safe halt during active external state modification, with evidence of state consistency preservation or rollback completion.

3. **Override demonstration:** empirical evidence that the human overseer can disregard, override, or reverse the system's output (Article 14(4)(d)), including evidence of output reversal capability for outputs that have propagated to downstream systems or external records.

4. **Oversight personnel competency:** documentation demonstrating that individuals assigned to human oversight possess the competency to understand the composed system's architecture, capabilities, and limitations, including the compositional risk categories applicable to the specific system.

5.3.6 Domain 6: Value Chain Agreement Evidence

Satisfies Article 25(4). Addresses all multi-vendor compositions.

1. **Article 25(4) agreements:** copies of or references to the written agreements with each third-party supplier of components, tools, services, or processes used or integrated in the composed system, demonstrating that the agreements specify the information, capabilities, technical access, and assistance required for compliance.

2. **Information sufficiency assessment:** a documented assessment of whether the information obtained from each upstream supplier under the Article 25(4) agreements is sufficient to support the compositional risk management (Domain 1), architecture documentation (Domain 2), and accuracy evaluation (Domain 4) evidence requirements. Where information is insufficient, the provider **SHALL** document the gap and the compensating measures applied.

3. **Update notification mechanisms:** documentation of the contractual and technical mechanisms through which upstream suppliers notify the system provider of component changes, enabling the provider to assess whether those changes constitute substantial modifications and to update post-market monitoring accordingly.

5.4 Internal Control Procedure (Annex VI) for Composed Systems

For Annex III high-risk systems where internal control applies, the provider **SHALL** execute the following adapted procedure:

1. **Quality management system verification:** verify that the QMS established under Article 17 addresses the compositional aspects of system development, deployment, and monitoring. The QMS **SHALL** include procedures for component integration testing, compositional risk assessment, upstream supplier management, and predetermined change governance.

2. **Evidence assembly:** assemble the evidence across all six domains defined in Section 5.3. The evidence **SHALL** be organised by domain and cross-referenced to the specific Chapter III requirements each evidence element satisfies.

3. **Compositional compliance verification:** for each Chapter III requirement, verify that compliance is demonstrated at the system level, not merely at the component level. The verification **SHALL** explicitly address the compositional implications identified in Section 2 for each requirement.

4. **EU declaration of conformity:** execute the EU declaration of conformity under Article 47, declaring that the composed system—as an integrated whole—complies with Chapter III, Section 2.

5. **Technical documentation filing:** ensure the Annex IV technical documentation, including the compositional evidence assembled under this methodology, is available to national competent authorities upon request and maintained for the period required by Article 18.

5.5 Third-Party Assessment Procedure (Annex VII) for Composed Systems

For systems requiring Annex VII assessment by a notified body (biometric identification systems under Article 43(1), Annex I product-embedded systems under Article 43(3)), the assessment **SHALL** include the following compositional elements in addition to the standard Annex VII procedure:

1. **Architecture review:** the notified body **SHALL** review the architecture documentation (Domain 2) and verify that the system boundary determination is complete, that all components within the boundary are identified, and that boundary characterisations are technically sound.

2. **Compositional risk review:** the notified body **SHALL** review the compositional risk register (Domain 1) and verify that the CRT taxonomy has been applied to the specific architecture, that all applicable risk categories have been assessed, and that mitigation measures are appropriate to the identified risks.

3. **Evidence audit:** the notified body **SHALL** audit the evidence across all six domains. Particular attention **SHALL** be given to the independence and representativeness of the system-level evaluation set used for accuracy evidence (Domain 4), the completeness of the adversarial testing programme relative to the compliance-critical boundaries identified in the architecture profile (Domain 4), the effectiveness of the safe halt mechanism verified through witnessed testing where feasible (Domain 5), and the sufficiency of Article 25(4) agreements relative to the information needs of the compositional assessment (Domain 6).

4. **Compositional testing:** where the notified body determines that the provider's evidence is insufficient, the body **MAY** conduct or commission independent testing of the composed system, including adversarial testing targeting compliance-critical boundaries and safe halt testing under operational conditions.

5. **Certificate scope:** the conformity certificate **SHALL** specify the system boundary at the time of assessment, the set of components within the boundary, and the predetermined change categories within which component modifications do not require re-assessment. Changes outside this scope invalidate the certificate under Article 43(4).

5.6 Substantial Modification Triggers for Composed Systems

Article 3(23) defines a substantial modification as a change not foreseen in the initial conformity assessment that affects compliance or modifies the intended purpose. For composed systems, the following change types **SHALL** be assessed against the substantial modification threshold.

Presumptively substantial (requires full re-assessment unless specific pre-documentation conditions are met):

1. Replacing a component model with a different model. May be non-substantial if pre-documented under Annex IV 2(f) and the replacement model falls within specified parametric boundaries (architecture, parameter count, training domain, accuracy range).

2. Adding a new MCP server or tool connection outside pre-documented categories. May be non-substantial if a matching tool category was pre-documented under Annex IV 2(f) and capability governance review is completed.

3. Adding a new agent to a multi-agent system. Generally constitutes a substantial modification as it alters the system's capability envelope. Pre-documentation of specific agent role categories may reduce classification to assessment-required.

4. Modifying the orchestration logic (routing algorithm structure, feedback loop architecture, delegation protocol). Pre-documented parametric ranges for specific parameters only; structural changes are substantial.

5. Changing the intended purpose or deploying the system in a new Annex III use case. Always substantial. Triggers full re-assessment and potentially reclassification under Article 6.

Assessment-required (must be evaluated against pre-documented parametric boundaries to determine whether substantial):

1. Updating a component model to a new version from the same provider. Non-substantial if pre-documented as a predetermined change, the update falls within the specified version range, and the provider confirms no material change to performance characteristics via Article 25(4) agreement.

2. Modifying routing thresholds or aggregation weights. Non-substantial if the modification falls within pre-documented parametric ranges and system-level accuracy remains within declared metrics.

3. Changing the retrieval corpus in a RAG pipeline. Non-substantial if corpus governance procedures are documented, the new corpus falls within the specified domain and quality parameters, and semantic alignment verification is completed.

5.7 Pre-Determination Strategy

The most effective mitigation of the substantial modification problem is a robust pre-determination strategy: systematically documenting, at the time of initial conformity assessment, the foreseeable component changes and their acceptable parametric boundaries under Annex IV 2(f). A well-executed pre-determination strategy transforms the substantial modification dilemma into a managed engineering discipline.

The pre-determination documentation **SHALL** include, for each foreseeable change category:

1. **Change description:** the type of component change anticipated (model version update, retrieval corpus refresh, tool addition within category, threshold adjustment).

2. **Parametric boundaries:** the measurable limits within which the change remains non-substantial. These boundaries **SHALL** be expressed in terms of system-level impact, not component-level specifications alone. For example: "replacement retrieval models shall achieve retrieval precision at $k = 10$ within $\pm 5\%$ of the assessed model's baseline on the reference evaluation set" rather than "replacement models shall have similar architecture."

3. **Verification procedure:** the test or measurement the provider will execute to confirm that a specific change falls within the pre-documented boundaries. This procedure **SHALL** be repeatable and auditable.

4. **Documentation update:** the documentation elements that will be updated to reflect the change, even when the change is non-substantial. Non-substantial changes still require documentation currency under Article 11.

5. **Escalation criteria:** the conditions under which the verification procedure's results indicate that the change exceeds the pre-documented boundaries, triggering escalation to a full substantial modification assessment.

The pre-determination strategy **SHALL** balance breadth and rigour. Pre-documenting excessively broad parametric boundaries provides operational flexibility but weakens the assessment's evidentiary value: an assessment that pre-approves all conceivable changes is functionally equivalent to no assessment. Excessively narrow boundaries provide strong evidentiary value but trigger frequent re-assessments that impede operational agility. The appropriate balance is architecture-dependent and risk-proportionate.

5.8 Minimum Upstream Documentation Requirements

Article 25(4) mandates written agreements specifying the information necessary for compliance. Article 53 requires GPAI providers to make information available to downstream system providers. For composed systems, the conformity assessment cannot proceed without minimum documentation from each upstream component supplier.

The system provider **SHALL** obtain, at minimum, the following information from each upstream supplier of a component integrated within the system boundary:

1. **Component capability specification:** the component's intended function, performance characteristics, known limitations, and operational envelope (the conditions under which the component is designed to operate correctly).
2. **Training data summary:** sufficient information about the component's training data to enable the system provider to assess data governance compliance under Article 10 and bias examination under Article 10(2)(f). For GPAI models, this is the summary specified in Annex XII.
3. **Accuracy and robustness metrics:** the component's declared accuracy and robustness characteristics, including the evaluation methodology and the conditions under which performance was measured.
4. **Known vulnerabilities and attack surfaces:** documented vulnerabilities, including adversarial input sensitivities, prompt injection susceptibility, and known failure modes relevant to the component's role in the composed system.
5. **Update notification commitment:** a contractual commitment to notify the system provider of material changes to the component, including model version updates, training data changes, and newly discovered vulnerabilities, within a timeframe sufficient to enable the system provider to assess whether the change constitutes a substantial modification.
6. **Cooperation commitment:** a commitment to cooperate with the system provider as required by Article 25(2)–(3), including reasonable technical access for integration testing and compositional safety assessment.

Where an upstream supplier declines to provide information in one or more categories, the system provider **SHALL** document the information gap and the compensating measures applied. Compensating measures **MAY** include independent testing of the component within the composed system, black-box adversarial evaluation, and increased post-market monitoring of the component's contribution to system-level performance. The information gap and compensating measures **SHALL** be disclosed in the conformity assessment evidence (Domain 6).

Appendix B provides a contract template that operationalises these minimum documentation requirements as Article 25(4) contractual provisions.

EU Compliance Series — Forthcoming

Specific parametric boundary templates for each architecture type, reference verification procedures, pre-determination strategy examples calibrated to different risk levels, and sector-specific upstream documentation requirement profiles will be provided in subsequent editions of the CRSA-1 EU Compliance Series.

6 Intersecting EU Regulatory Obligations

6.1 The Compound Compliance Problem

Composed AI systems deployed in the European Union do not operate under the AI Act alone. Seven intersecting EU regulations impose obligations that compound the compositional safety challenge. Each regulation was drafted independently, for a different regulatory purpose, and none contemplates the compositional architectures that now dominate production AI. The result is a compliance environment where a single composed system may simultaneously trigger obligations under the AI Act, DORA, MDR, NIS2, GDPR, the Product Liability Directive, and the Cyber Resilience Act—with no harmonized methodology for satisfying all obligations concurrently.

The Digital Omnibus on AI (COM(2025) 836) proposes a single-entry-point mechanism for breach notifications to reduce reporting duplication across NIS2, GDPR, and the AI Act. However, this addresses procedural overlap, not substantive overlap. The underlying obligations remain distinct, cumulative, and compositionally blind.

This section maps each intersecting regulation to its specific compositional implications, identifying where the compound obligations exceed what any single regulatory compliance programme addresses.

6.2 DORA — Digital Operational Resilience Act

Regulation 2022/2554 — Fully Applicable Since 17 January 2025

Sector: Financial services (banks, insurance undertakings, investment firms, payment institutions).

Core Obligation	Financial entities must implement ICT risk management frameworks, manage third-party ICT service provider risk, and conduct digital operational resilience testing (Articles 5–16, 28–30).
Compositional Implication	Each AI model provider in a composed pipeline must be treated as an ICT third-party service provider under Article 28. Article 28(3) mandates a Register of Information documenting all contractual arrangements with AI providers. Article 29 requires ICT concentration risk assessment—directly relevant when multiple models in a pipeline originate from the same foundation model vendor. Article 30 mandates contractual provisions including SLAs, audit rights, and exit strategies for each AI component provider. A bank deploying a multi-model credit scoring pipeline must comply with both AI Act Article 9 risk management and DORA's ICT risk framework. Neither addresses emergent risks from model composition. The DORA Register of Information must document each model provider; the AI Act Article 25(4) agreement must specify technical information flows. These are parallel contractual obligations with different scopes, different enforcement authorities, and no harmonized template.
CRSA-1 Mechanism	The Article 25(4) contract template in Appendix B is designed to be compatible with DORA Article 30 contractual requirements, enabling a single agreement to satisfy both regulatory frameworks. The compositional risk register (Section 5.3, Domain 1) provides the risk assessment infrastructure that both Article 9 and DORA Article 6 require.

6.3 MDR — Medical Devices Regulation

Regulation 2017/745 — Applicable

Sector: Healthcare (medical devices, Software as a Medical Device, in vitro diagnostics under IVDR 2017/746).

Core Obligation	AI systems functioning as Software as a Medical Device (SaMD) must undergo clinical evaluation and conformity assessment under MDR. Classification follows Annex VIII Rule 11.
Compositional Implication	When a medical AI system composes multiple models—for example, a vision model for radiological image analysis combined with an LLM for clinical report synthesis—the overall system's intended use determines MDR classification and the clinical evaluation must prove the clinical safety of the composed output, not merely of individual components. AI Act Article 6(1) automatically classifies medical AI requiring notified body MDR assessment as high-risk AI. The MDCG 2025-6 guidance (June 2025) on MDR/AI Act interaction clarifies that the MDR notified body performs both MDR and AI Act assessment for the integrated product. However, when the medical device integrates a third-party GPAI model, the manufacturer must ensure entire system safety but may lack full access to the GPAI model's internals—a structural tension the guidance acknowledges but does not resolve. Composed medical AI systems face dual conformity assessment: clinical safety under MDR and compositional compliance under the AI Act. The notified body must evaluate both the clinical performance of the composed output and the compositional architecture through which that output is produced. No notified body has published a methodology for this dual assessment.
CRSA-1 Mechanism	The agentic system compliance profile (Section 4.3) and the conformity assessment methodology (Section 5) provide the compositional assessment framework that MDR notified bodies require for evaluating multi-model medical AI systems.

6.4 NIS2 — Network and Information Security Directive

Directive 2022/2555 — Transposition Required by 17 October 2024

Sector: Essential and important entities across energy, transport, banking, health, digital infrastructure, and ICT service management.

Core Obligation	Essential and important entities must implement supply chain security measures (Article 21(2)(d)) and report significant incidents within 24 hours (early warning) and 72 hours (full notification) under Article 23.
Compositional Implication	For organisations deploying multi-vendor AI, NIS2 requires including each AI component supplier in supply chain risk assessments and evaluating each vendor's cybersecurity practices. Incidents propagating through AI pipelines—such as a context poisoning attack (CRT-6) that traverses from a retrieval corpus through a generation model to produce harmful output—must be detected and reported within the 24-hour early warning window. Definitively establishing which component in a composed system failed, to trigger the correct NIS2 notification, requires the cross-component logging infrastructure specified in Section 4 for each architecture type. Without compositional logging, the entity cannot attribute the incident to a specific supply chain element within the reporting timeline.
CRSA-1 Mechanism	The logging architecture requirements in Section 4 provide the cross-component traceability necessary for NIS2 incident attribution. The compositional cybersecurity assessment (Section 5.3, Domain 4) addresses the supply chain security evaluation NIS2 Article 21(2)(d) requires for AI components.

6.5 GDPR — General Data Protection Regulation

Regulation 2016/679 — Applicable

Sector: All sectors processing personal data of individuals in the EU.

Core Obligation	Controllers must ensure lawful processing, implement data minimisation, conduct Data Protection Impact Assessments (DPIAs) for high-risk automated processing (Article 35), and provide meaningful information about automated decision-making logic (Articles 13(2)(f), 15(1)(h), and 22).
Compositional Implication	Multi-model pipelines inherently pass personal data between components—and potentially between independent third-party APIs—triggering complex controller/processor relationships under Articles 26 and 28. Each model provider's role in the processing chain must be defined: is the GPAI provider a processor, a joint controller, or an independent controller for the data it processes? The answer determines data protection responsibilities and contractual obligations. Article 22(1) regulates automated decision-making. For multi-model pipelines, the fact that decisions pass through multiple models does not change the "solely automated" character under EDPB interpretation. Articles 13(2)(f) and 15(1)(h) require "meaningful information about the logic involved"—substantially harder when the decision flows through multiple opaque models. The DPIA required by Article 35(3)(a) must assess the entire pipeline, not individual components. The AI Act Article 9(4) interaction effects obligation creates a specific GDPR tension for composed systems: the logging granularity required by AI Act Article 12 for compositional traceability may conflict with GDPR data minimisation principles. Recording the full processing path for every invocation—including all retrieved documents, intermediate model outputs, and agent reasoning chains—creates extensive personal data records that must themselves be governed under GDPR.
CRSA-1 Mechanism	The interaction effects assessment (Section 5.3, Domain 1) is designed to capture the AI Act/GDPR logging-versus-minimisation tension. The value chain agreement template (Appendix B) includes provisions for GDPR controller/processor determination alongside Article 25(4) information-sharing obligations.

6.6 Product Liability Directive

Directive 2024/2853 — Transposition by 9 December 2026

Sector: All sectors placing products (including AI systems and their components) on the EU market.

Core Obligation	Manufacturers and integrators face strict liability for defective products. The Directive explicitly includes AI systems and their components as "products" subject to its provisions.
Compositional Implication	Three provisions create acute compositional liability exposure. Article 8(1)(a) provides that the integrator is liable for defective component models "integrated into, or inter-connected with, a product within that manufacturer's control." The entity assembling a composed AI system bears strict liability for defects in any integrated component. Article 12 establishes joint and several liability where multiple operators are liable for the same damage. When a composed AI system causes harm and the defect cannot be attributed to a single component—the paradigmatic compositional failure scenario—all operators in the value chain may be jointly liable. Most significantly, Article 10(2)(b) creates a rebuttable presumption of defectiveness when the claimant proves non-compliance with mandatory safety requirements, including the AI Act. For composed systems, failure to comply with any of the Chapter III requirements mapped in Section 2—because no compositional safety methodology was available—triggers a presumption that the system was defective. Article 10(4) further presumes defectiveness when proof is "excessively difficult due to technical or scientific complexity"—a provision explicitly designed for AI. The combined effect is severe: AI Act non-compliance for a composed system creates a presumption of defectiveness under the Product Liability Directive, shifting the burden of proof to the provider. Compositional safety compliance is no longer merely a regulatory obligation; it is a liability shield.
CRSA-1 Mechanism	The conformity assessment methodology (Section 5) provides the compliance evidence that rebuts the presumption of defectiveness. Documented compositional risk management, architecture documentation, and system-level accuracy evidence directly address the Article 10(2)(b) trigger.

6.7 Cyber Resilience Act

Regulation 2024/2847 — Full Applicability from 11 December 2027

Sector: All manufacturers of products with digital elements placed on the EU market.

Core Obligation	Manufacturers must exercise cybersecurity due diligence when integrating third-party components, ensure vulnerability handling throughout the product lifecycle, and report actively exploited vulnerabilities within 24 hours.
Compositional Implication	Article 13(5) requires manufacturers integrating third-party components to "exercise due diligence when integrating components sourced from third parties so that those components do not compromise the cybersecurity of the product." For AI model providers, this extends cybersecurity due diligence to every component in the composed pipeline. Actively exploited vulnerabilities in AI components—including adversarial attacks, prompt injection vectors, and model extraction techniques—must be reported within 24 hours. For composed systems, a vulnerability in one component that is only exploitable through the compositional architecture (e.g., an indirect prompt injection that requires the retrieval-to-execution chain to succeed) creates an attribution challenge: is this a vulnerability in the component or in the composition? Article 12 of the CRA provides that high-risk AI systems meeting CRA Annex I cybersecurity requirements are deemed compliant with AI Act Article 15 cybersecurity requirements. However, AI Act accuracy and robustness requirements remain separate—CRA compliance does not satisfy the full Article 15 obligation for composed systems.
CRSA-1 Mechanism	The compositional cybersecurity assessment (Section 5.3, Domain 4) and the CRT-6 Context Poisoning risk category (Section 3) address the due diligence requirement for compositional cybersecurity.

6.8 AI Liability Directive — Withdrawn

COM(2022)496 — Formally Withdrawn October 2025

Status	The AI Liability Directive proposal was formally withdrawn by the Commission in October 2025 following its decision of 16 July 2025. There is now no EU-level harmonization of fault-based AI liability.
Compositional Implication	Fault-based claims for negligent composition of AI systems—for example, negligent selection of component models, negligent integration testing, or negligent monitoring of compositional health—revert to national tort law across 27 Member States. This creates potential fragmentation: the standard of care for composing AI systems may vary by jurisdiction. The Product Liability Directive's strict liability regime remains the only harmonized instrument. Combined with its presumption of defectiveness for AI Act non-compliance (Article 10(2)(b)), this creates a regulatory environment where strict liability (via PLD) is harmonized but fault-based liability (via national tort law) is fragmented. For composed system providers, this means that demonstrating AI Act compliance through the conformity assessment methodology in Section 5 provides a harmonized liability defence under the PLD, but does not necessarily address national fault-based claims.

6.9 Compound Obligation Summary

The following table summarises the intersecting regulatory obligations by sector, illustrating the compound compliance burden for composed AI systems in regulated industries.

Sector	Applicable Regulations	Earliest Enforcement
Financial Services	AI Act + DORA + GDPR + NIS2 + PLD + CRA	DORA: Jan 2025 (applicable). AI Act: Aug 2026.
Healthcare / MedTech	AI Act + MDR/IVDR + GDPR + NIS2 + PLD + CRA	MDR: applicable. AI Act: Aug 2026/2027.
Critical Infrastructure	AI Act + NIS2 + GDPR + PLD + CRA	NIS2: Oct 2024 (transposition). AI Act: Aug 2026.
Employment / HR	AI Act + GDPR + PLD	AI Act: Aug 2026. GDPR: applicable.
Government Services	AI Act + GDPR + NIS2 + PLD	AI Act: Aug 2026. NIS2: Oct 2024.
Insurance	AI Act + DORA + GDPR + PLD	DORA: Jan 2025. AI Act: Aug 2026.

In every regulated sector, composed AI systems face a minimum of three intersecting regulatory frameworks. Financial services and healthcare face six simultaneous frameworks. No exist-

ing compliance methodology addresses the compound obligation across these frameworks for composed systems.

EU Compliance Series — Forthcoming

Sector-specific editions of the CRSA-1 EU Compliance Series will provide integrated compliance profiles addressing the compound obligations for each sector. The Financial Services Edition will address AI Act and DORA dual-compliance. The Medical Devices Edition will address AI Act and MDR/IVDR dual-conformity assessment. The Critical Infrastructure Edition will address AI Act and NIS2 supply chain security requirements. Each edition will provide sector-specific Article 25(4) contract templates, risk register extensions, and conformity assessment adaptations.

7 Implementation Roadmap

7.1 Purpose

This section provides a phased implementation timeline for organisations deploying composed AI systems classified as high-risk under the AI Act. The roadmap is structured around two enforcement scenarios: the baseline scenario under the enacted AI Act text (enforcement from 2 August 2026) and the contingency scenario under the Digital Omnibus proposal (enforcement from the Commission readiness decision, with long-stop dates of 2 December 2027 for Annex III and 2 August 2028 for Annex I).

Regardless of scenario, the Product Liability Directive transposition deadline of 9 December 2026 means that AI Act non-compliance triggers a presumption of defectiveness from that date forward. Organisations that defer compositional safety compliance pending the Digital Omnibus outcome accept liability exposure from December 2026 even if AI Act enforcement is delayed.

7.2 Phase 1: Foundation (Now Through June 2026)

Objective: Establish compositional awareness and documentation infrastructure.

1. **System inventory and classification.** Identify all AI systems in deployment or development that meet the composed system definition in Section 1.2. For each system, determine whether it falls within Annex I or Annex III high-risk classification under Article 6. Apply the Article 6(3) derogation analysis at the system level, not the component level.

2. **Provider determination.** For each composed system, apply the provider determination criteria in Section 5.2.2 to identify the entity bearing Article 16 obligations. Where the organisation is a deployer that has integrated or modified components, assess whether Article 25(1) reclassification applies.

3. **Architecture documentation.** Begin documenting each composed system's architecture per the applicable Section 4 profile and the Annex IV 2(c) requirements. Identify all components within the system boundary, characterise all compliance-critical boundaries, and document the data flow architecture.

4. **Compositional risk assessment.** Apply the CRT taxonomy (Section 3) to each composed system. Produce the initial compositional risk register required by Section 5.3, Domain 1. Identify the highest-severity risk categories for each architecture and prioritise mitigation.

5. **Value chain mapping.** Identify all third-party component suppliers for each composed system. Assess the current state of information sharing against the minimum upstream

documentation requirements in Section 5.8. Initiate Article 25(4) agreement negotiations using the template in Appendix B.

6. **QMS extension.** Extend the organisation's quality management system (Article 17) to include compositional system management procedures: component integration testing, compositional risk assessment, upstream supplier management, and predetermined change governance.

7.3 Phase 2: Compliance Infrastructure (June Through December 2026)

Objective: Deploy the technical infrastructure required for ongoing compositional safety compliance.

1. **Logging deployment.** Implement the architecture-specific logging requirements from Section 4 for each composed system. Verify cross-component correlation capability through sample trace reconstructions. Ensure logging granularity satisfies both AI Act Article 12 and NIS2 incident attribution requirements.

2. **Human oversight implementation.** Deploy the oversight architecture specified in the applicable Section 4 profile, including intervention points, safe halt mechanisms, and override capabilities. For agentic and MCP-based systems, implement and test the safe halt procedure including state rollback.

3. **System-level evaluation.** Conduct system-level accuracy and robustness evaluation per the Article 15 requirements in the applicable Section 4 profile. Produce the compositional accuracy and robustness evidence required by Section 5.3, Domain 4. Declare accuracy metrics in the instructions of use.

4. **Adversarial testing.** Conduct compositional adversarial testing targeting each compliance-critical boundary identified in the architecture profile. Include the boundary-specific adversarial scenarios specified in Section 4. Document results as part of the Domain 4 evidence package.

5. **Pre-determination documentation.** Prepare the predetermined change specification under Annex IV 2(f), applying the pre-determination strategy in Section 5.7. Define parametric boundaries, verification procedures, and escalation criteria for each foreseeable component change category.

6. **FRIA execution.** For systems subject to Article 27, conduct the fundamental rights impact assessment incorporating compositional bias analysis. Evaluate routing discrimination (CRT-4) and aggregation bias (CRT-3) at the system level.

7. **Product Liability Directive preparation.** Ensure that the conformity assessment evidence assembled under Section 5 is sufficient to rebut the presumption of defectiveness under PLD Article 10(2)(b) by the 9 December 2026 transposition deadline.

7.4 Phase 3: Conformity and Continuous Compliance (From August 2026 or Omnibus Trigger)

Objective: Execute conformity assessment and establish continuous compliance monitoring.

1. **Conformity assessment execution.** Assemble the evidence across all six domains (Section 5.3). For Annex VI internal control systems, execute the adapted procedure in Section 5.4. For Annex VII systems, engage a notified body and provide the compositional evidence package for third-party assessment per Section 5.5.

2. **EU declaration and registration.** Execute the EU declaration of conformity (Article 47) and register the composed system in the EU database (Article 49).

3. **Post-market monitoring activation.** Deploy the post-market monitoring system required by Articles 61 and 72, incorporating compositional health metrics: inter-component drift detection, aggregation stability monitoring, capability inventory verification, and cross-component performance correlation tracking.

4. **Continuous compositional monitoring.** Establish ongoing monitoring for each applicable CRT category. Monitor for cascade failure signatures, semantic drift indicators, aggregation bias emergence, routing fairness metrics, capability inventory changes, context poisoning indicators, and state consistency verification results.

5. **Change management.** Operate the predetermined change governance framework: for each component change, execute the verification procedure to determine whether the change falls within pre-documented parametric boundaries or requires escalation to a substantial modification assessment. Document all changes and verification results regardless of classification.

6. **Incident response.** Establish incident response procedures that satisfy AI Act serious incident reporting (Article 73), NIS2 24-hour early warning requirements (Article 23), and DORA incident classification (where applicable), with compositional logging providing the attribution evidence necessary for each reporting obligation.

7.5 Timeline Visualisation

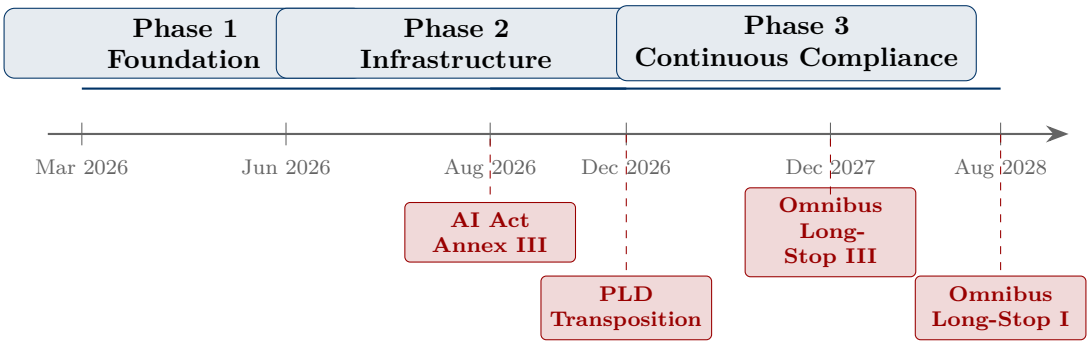

Figure 7: Implementation roadmap timeline. Phase 1 (foundation) through June 2026. Phase 2 (infrastructure) through December 2026. Phase 3 (continuous compliance) from August 2026 onward. Red milestones indicate enforcement dates and liability triggers. Phases 2 and 3 overlap intentionally: infrastructure deployment continues while conformity assessment proceeds.

7.6 Digital Omnibus Contingency

If the Digital Omnibus on AI is adopted before 2 August 2026, the following adjustments apply:

1. **Phase 2 timeline extension.** The compliance infrastructure deployment window extends from the Commission readiness decision (unknown date) through six months after that decision for Annex III systems, with 2 December 2027 as the absolute long-stop.

2. **Phase 1 urgency unchanged.** The foundation phase is not affected by the Omnibus. Architecture documentation, compositional risk assessment, and value chain mapping should proceed regardless of enforcement timeline, because the Product Liability Directive transposition deadline of 9 December 2026 creates liability exposure independent of AI Act enforcement.

3. **Pre-determination advantage.** Organisations that complete Phase 1 and begin Phase 2 before the Omnibus outcome is known gain a strategic advantage: their predetermined change documentation will be in place before enforcement begins, enabling operational agility from the first day of compliance.

The Omnibus, if adopted, provides additional time but does not reduce the scope of compliance obligations. Every requirement mapped in Section 2 applies in full once the enforcement trigger activates. The long-stop dates are not deferrals; they are deadlines.

A CRSA-1 / EU AI Act Cross-Reference Matrix

This appendix provides a consolidated cross-reference mapping every CRSA-1 element to its corresponding AI Act provision. The matrix is designed for use by compliance teams, auditors, and notified bodies as a verification checklist.

A.1 Section-to-Article Mapping

CRSA-1 Section	AI Act Provision	Compliance Function
Section 2 — Obligation Mapping	Articles 3, 6, 8–15, 25–27, 43, 51–53, 61, 72, 86	Identifies compositional implications of each obligation
Section 3 — Risk Taxonomy (CRT-1 to CRT-8)	Article 9(1)–(2), 9(4)	Provides the risk classification framework for compositional risks
Section 4 — RAG Profile	Articles 9, 11, 12, 13, 14, 15	Architecture-specific compliance requirements for RAG pipelines
Section 4 — Agentic Profile	Articles 9, 11, 12, 14, 15	Architecture-specific compliance requirements for agentic systems
Section 4 — Ensemble Profile	Articles 9, 10, 11, 12, 13, 14, 15	Architecture-specific compliance requirements for ensemble systems
Section 4 — Cascade Profile	Articles 9, 11, 12, 15, 27	Architecture-specific compliance requirements for routing systems
Section 4 — MCP Multi-Agent Profile	Articles 9, 11, 12, 14, 15, 43(4)	Architecture-specific compliance requirements for multi-agent systems
Section 5.2 — System Boundary	Annex IV 1(b)	Defines the composed system boundary for assessment scope
Section 5.2 — Provider Determination	Articles 3(3), 25(1)	Determines the responsible provider for the composed system
Section 5.3 — Domain 1: Risk Management	Article 9	Evidence requirements for compositional risk management

CRSA-1 Section	AI Act Provision	Compliance Function
Section 5.3 — Domain 2: Architecture	Article 11, Annex IV 2(a)(c)(f)	Evidence requirements for architecture documentation
Section 5.3 — Domain 3: Logging	Article 12	Evidence requirements for cross-component traceability
Section 5.3 — Domain 4: Accuracy	Article 15	Evidence requirements for system-level accuracy and robustness
Section 5.3 — Domain 5: Oversight	Article 14	Evidence requirements for human oversight and safe interruption
Section 5.3 — Domain 6: Value Chain	Article 25(4)	Evidence requirements for multi-vendor agreements
Section 5.4 — Internal Control	Annex VI	Adapted Annex VI procedure for composed systems
Section 5.5 — Third-Party Assessment	Annex VII	Adapted Annex VII procedure for composed systems
Section 5.6 — Substantial Modification	Article 3(23), 43(4)	Modification classification for composed systems
Section 5.7 — Pre-Determination	Annex IV 2(f)	Strategy for predetermined change documentation
Section 5.8 — Upstream Documentation	Articles 25(4), 53	Minimum documentation from component suppliers
Section 6 — Intersecting Regulations	(Cross-regulatory)	Compound compliance mapping for DORA, MDR, NIS2, GDPR, PLD, CRA
Section 7 — Implementation Roadmap	Articles 43, 47, 49, 61, 72	Phased implementation timeline with enforcement milestones
Appendix B — Contract Template	Article 25(4)	Model contractual terms for multi-vendor composed systems

A.2 CRT Category to Article Mapping

CRT Category	Primary Articles	Evidence Domain
CRT-1 Cascade Failure	Art. 9, 12, 15(1), 15(3)	Domain 1 (Risk), Domain 3 (Logging), Domain 4 (Accuracy)
CRT-2 Semantic Drift	Art. 10, 11, 15(1), 15(2)	Domain 2 (Architecture), Domain 4 (Accuracy)

CRT Category	Primary Articles	Evidence Domain
CRT-3 Aggregation Bias	Art. 9, 10(2)(f), 27, 86	Domain 1 (Risk), Domain 4 (Accuracy)
CRT-4 Routing Discrimination	Art. 9(4), 15(1), 27, 86	Domain 1 (Risk), Domain 4 (Accuracy)
CRT-5 Capability Drift	Art. 3(23), 11, 13, 43(4), 51	Domain 2 (Architecture), Domain 1 (Risk)
CRT-6 Context Poisoning	Art. 9(8), 12, 15(4)–(5), 25(4)	Domain 4 (Accuracy), Domain 6 (Value Chain)
CRT-7 State Corruption	Art. 12, 14(4), 15(3), 61	Domain 3 (Logging), Domain 5 (Oversight)
CRT-8 Compositional Opacity	Art. 13, 14(4)(a), 86	Domain 2 (Architecture), Domain 5 (Oversight)

B Article 25(4) Model Contract Terms for Composed AI Systems

B.1 Preamble

Article 25(4) of Regulation (EU) 2024/1689 mandates that the provider of a high-risk AI system and any third party supplying AI systems, tools, services, components, or processes used or integrated in that system shall, by written agreement, specify the necessary information, capabilities, technical access, and other assistance required for the provider to comply with its obligations under the Regulation.

The AI Office is empowered to develop voluntary model contractual terms but has not published any as of March 2026. This Appendix provides the first published model contract terms for composed AI systems. These terms are designed to satisfy the Article 25(4) mandate while remaining compatible with DORA Article 30 contractual requirements for financial entities and GDPR Article 28 data processing agreements.

These terms are provided as a template. Organisations should adapt them to their specific circumstances, component relationships, and sectoral regulatory requirements. They are not legal advice.

Honest Framing

These model contract terms provide a structural framework for Article 25(4) compliance in multi-vendor AI systems. They do not constitute legal advice, do not guarantee regulatory approval, and do not substitute for review by qualified legal counsel familiar with the specific jurisdictional and sectoral context of the deployment. The terms address the compositional safety dimensions of the contractual relationship; they do not address the full scope of commercial, intellectual property, or liability provisions that a comprehensive AI component supply agreement would require.

B.2 Definitions

For the purposes of these model terms:

1. **System Provider** means the entity identified as the provider of the high-risk AI system under Articles 3(3) or 25(1) of the Regulation.
2. **Component Supplier** means the third party supplying an AI system, tool, service, component, or process that is used or integrated in the System Provider's high-risk AI system.
3. **Component** means the specific AI system, model, tool, service, or process supplied by the Component Supplier and integrated into the System Provider's composed system.
4. **System Boundary** means the set of components, data sources, tools, and interfaces that constitute the AI system being assessed, as defined by the System Provider under Section 5.2 of the CRSA-1 specification.
5. **Material Change** means any modification to the Component that alters its performance characteristics, training data, capability envelope, or known limitations beyond the ranges documented in the Component Specification.

B.3 Information Provisions

Clause B.1 — Component Specification Delivery

The Component Supplier **SHALL** provide to the System Provider, within [specified timeframe] of contract execution, a Component Specification containing, at minimum:

(a) the Component's intended function, performance characteristics, and operational envelope;
(b) a summary of the Component's training data sufficient to enable the System Provider to assess compliance with Article 10 of the Regulation;
(c) the Component's declared accuracy and robustness metrics, including the evaluation methodology and conditions under which performance was measured;
(d) known vulnerabilities, adversarial input sensitivities, and failure modes relevant to the Component's role in the composed system;
(e) known limitations on the Component's performance, including input domains, data distributions, or operational conditions under which performance degrades; and
(f) the Component's version identifier and release date.

Clause B.2 — Update Notification

The Component Supplier **SHALL** notify the System Provider of any Material Change to the Component no fewer than [specified number] business days before the change takes effect. The notification **SHALL** include:

(a) a description of the change;
(b) the anticipated impact on the Component's performance characteristics, training data, capability envelope, or known limitations;
(c) an updated Component Specification reflecting the change; and
(d) an assessment of whether the change affects the System Provider's ability to comply with the Regulation.

Where immediate deployment of a change is required for security or safety reasons, the Component Supplier **SHALL** notify the System Provider as soon as reasonably practicable and no later than [specified number] business days after deployment, with the information specified above.

Clause B.3 — Technical Access for Integration Testing

The Component Supplier **SHALL** provide the System Provider with reasonable technical access to the Component sufficient to enable:

(a) system-level integration testing of the Component within the composed system;
(b) compositional accuracy and robustness evaluation at the system level;
(c) adversarial testing targeting the compliance-critical boundary between the Component and adjacent components in the composed system; and
(d) verification that the Component operates within its documented performance envelope when deployed in the specific composed architecture.

Technical access **SHALL** be provided in a manner that protects the Component Supplier's trade secrets and proprietary information, in accordance with Article 78 of the Regulation.

Clause B.4 — Incident Cooperation

In the event of a serious incident (as defined in Article 3(49) of the Regulation) involving the composed system, the Component Supplier **SHALL**:

(a) cooperate with the System Provider in the investigation of the incident, including providing diagnostic information about the Component's behaviour during the incident period;
(b) provide reasonable technical assistance to enable the System Provider to determine whether the Component contributed to the incident; and
(c) cooperate with the System Provider in the preparation of any serious incident report required under Article 73 of the Regulation.

Cooperation obligations under this clause are subject to the "reasonably expected" standard of Article 25(2) and do not require the Component Supplier to disclose proprietary model internals beyond what is necessary for incident investigation.

Clause B.5 — Ongoing Cooperation

The Component Supplier **SHALL** cooperate with the System Provider as required by Article 25(2)–(3) of the Regulation, including:

(a) providing reasonable assistance for the System Provider's conformity assessment (Annex VI or Annex VII) to the extent that the assessment requires information about or access to the Component;
(b) cooperating with national competent authority requests for information about the Component, where such requests are channelled through the System Provider; and
(c) responding to reasonable requests for information from the System Provider necessary for the System Provider's post-market monitoring obligations under Articles 61 and 72.

Clause B.6 — DORA Compatibility (Financial Services)

Where the System Provider is a financial entity subject to Regulation (EU) 2022/2554 (DORA) and the Component Supplier is an ICT third-party service provider:

(a) the Component Supplier **SHALL** support the System Provider's Register of Information obligations under DORA Article 28(3) by providing information about the Component sufficient for inclusion in the register;

(b) the Component Supplier **SHALL** cooperate with the System Provider's digital operational resilience testing under DORA Articles 24–27 to the extent that testing involves the Component; and

(c) the provisions of this agreement are intended to be compatible with and supplementary to the contractual requirements of DORA Article 30.

This clause applies only where both the System Provider and the Component Supplier are subject to DORA or have agreed to DORA-aligned contractual provisions.

Clause B.7 — Data Protection

Where the Component processes personal data as part of the composed system:

(a) the parties **SHALL** determine and document their respective roles as controller, joint controller, or processor under Regulation (EU) 2016/679 (GDPR) with respect to personal data processed by the Component within the composed system;

(b) where the Component Supplier acts as a processor, a data processing agreement satisfying GDPR Article 28(3) **SHALL** be executed as a supplement to or incorporated within this agreement; and

(c) the Component Supplier **SHALL** provide the System Provider with sufficient information about the Component's data processing activities to enable the System Provider to complete any Data Protection Impact Assessment required under GDPR Article 35 for the composed system.

B.4 Template Usage Notes

These model terms are designed to be incorporated into a broader component supply agreement. They address the Article 25(4) compositional safety obligations specifically and do not replace standard commercial terms covering pricing, service levels, liability limitations, governing law, or dispute resolution.

Organisations adapting these terms should:

1. Specify the timeframes indicated by brackets (notification periods, specification delivery deadlines) based on the operational tempo of their specific deployment.

2. Review the technical access provisions (Clause B.3) with the Component Supplier to ensure compatibility with the supplier's trade secret protection policies and the Article 78 confidentiality framework.

3. Include Clause B.6 only where DORA applies and remove or adapt it for non-financial-services deployments.

4. Include Clause B.7 only where personal data processing is involved and ensure alignment with any existing GDPR data processing agreements.

5. Seek legal review by counsel qualified in both AI regulation and the applicable sectoral regulatory framework.

C Glossary of Compositional Safety Terms

Aggregation Bias (CRT-3)
An emergent discriminatory outcome produced by combining individually compliant AI components through a voting, weighting, or averaging mechanism, where the bias is a property of the aggregation architecture rather than of any individual component. See Section 3.3.3.

Capability Drift (CRT-5)
The post-deployment change in a composed system's functional capabilities through the addition, removal, or modification of connected tools, data sources, or component models, without corresponding conformity assessment update. See Section 3.3.5.

Cascade Failure (CRT-1)
The propagation and amplification of an error through sequential components in a composed system, where each component performs correctly given its input but the chain of correct local operations produces an incorrect global output. See Section 3.3.1.

Compliance-Critical Boundary
An interface between components in a composed AI system where data transformation, semantic interpretation, or trust delegation occurs and where compositional failures are most likely to originate. Identified in the architecture reference diagrams in Section 4.

Composed AI System
An AI system integrating two or more AI components—including but not limited to foundation models, specialised models, retrieval systems, routing classifiers, and tool-calling interfaces—into a pipeline, ensemble, or multi-agent architecture whose system-level behaviour emerges from the interaction of its components.

Compositional Opacity (CRT-8)
The condition where the interaction between components in a composed AI system produces decision processes that cannot be meaningfully explained, attributed, or decomposed into component contributions, even when individual components are independently interpretable. See Section 3.3.8.

Compositional Risk Taxonomy (CRT)
The eight-category risk classification framework established in Section 3, providing the vocabulary and structure for identifying, assessing, and managing risks specific to composed AI systems.

Context Poisoning (CRT-6)
The introduction of adversarial content into the data environment from which a composed AI system retrieves or receives context, causing downstream components to act on malicious instructions or corrupted information. The compositional form of indirect prompt injection. See Section 3.3.6.

Evidence Domain
One of six categories of conformity assessment evidence defined in Section 5.3, each mapped to specific Chapter III requirements: (1) Risk Management, (2) Architecture Documentation, (3) Logging and Traceability, (4) Accuracy and Robustness, (5) Human Oversight, (6) Value Chain Agreements.

Predetermined Change
A modification to a high-risk AI system that has been pre-documented by the provider and assessed at the time of initial conformity assessment under Annex IV 2(f), such that

the modification does not require a new conformity assessment under Article 43(4). For composed systems, the pre-determination strategy (Section 5.7) defines the parametric boundaries within which component changes qualify as predetermined.

Provider Determination
The process of identifying, under Articles 3(3) and 25, the entity bearing Article 16 provider obligations for a composed AI system. See Section 5.2.2.

Routing Discrimination (CRT-4)
Systematically differential treatment of demographic groups produced by a classifier, router, or threshold mechanism that directs inputs to processing pathways of different quality, where the discrimination arises from the routing architecture's interaction with downstream pathway quality differentials. See Section 3.3.4.

Semantic Drift (CRT-2)
The alteration of meaning or representational content as data passes between components operating in different embedding spaces, tokenisation schemes, or internal representations. See Section 3.3.2.

State Corruption (CRT-7)
Inconsistent or partially updated shared state resulting from the interruption, failure, or timeout of one component in a composed system, where downstream components or subsequent operations inherit the corrupted state. See Section 3.3.7.

Substantial Modification
A change to a high-risk AI system after it has been placed on the market or put into service which is not foreseen or planned in the initial conformity assessment and as a result of which the compliance of the system with the requirements of the Regulation is affected or the intended purpose is modified (Article 3(23)). For composed systems, Section 5.6 provides classification guidance for common change types.

System Boundary
The set of components, data sources, tools, and interfaces that constitute the AI system being assessed under conformity assessment. Defined using the criteria in Section 5.2.1. Components within the boundary are covered by the conformity assessment; systems at or beyond the boundary are documented under Annex IV 1(b) but assessed as external interactions.

References

Primary Legislation

[1] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *OJ L, 2024/1689, 12.7.2024.*

[2] Regulation (EU) 2022/2554 of the European Parliament and of the Council of 14 December 2022 on digital operational resilience for the financial sector (DORA). *OJ L 333, 27.12.2022.*

[3] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices (MDR). *OJ L 117, 5.5.2017.*

[4] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union (NIS2). *OJ L 333, 27.12.2022.*

[5] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data (GDPR). *OJ L 119, 4.5.2016.*

[6] Directive (EU) 2024/2853 of the European Parliament and of the Council of 23 October 2024 on liability for defective products (Product Liability Directive). *OJ L, 2024/2853, 18.11.2024.*

[7] Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements (Cyber Resilience Act). *OJ L, 2024/2847, 20.11.2024.*

Commission Proposals and Guidance

[8] European Commission. Proposal for a Regulation amending Regulations (EU) 2024/1689 and (EU) 2024/2847 (Digital Omnibus on AI). COM(2025) 836 final, 6 February 2025.

[9] European Commission. Commission Guidelines on the definition of an artificial intelligence system (C(2025) 3034 final), 4 February 2025.

[10] European Commission. Commission Guidelines on prohibited artificial intelligence practices (C(2025) 883 final), 4 February 2025.

Standards and Technical Specifications

[11] CEN-CENELEC JTC 21. prEN 18228: Artificial intelligence — Risk management. Under development.

[12] CEN-CENELEC JTC 21. prEN 18229-1: Artificial intelligence — Transparency — Part 1: General requirements. Under development.

[13] CEN-CENELEC JTC 21. prEN 18229-2: Artificial intelligence — Transparency — Part 2: AI system testing and validation. Under development.

[14] CEN-CENELEC JTC 21. prEN 18282: Artificial intelligence — Cybersecurity requirements. Under development.

[15] CEN-CENELEC JTC 21. prEN 18284: Artificial intelligence — Data quality. Under development.

[16] CEN-CENELEC JTC 21. prEN 18285: Artificial intelligence — Conformity assessment. Under development.

[17] CEN-CENELEC JTC 21. prEN 18286: Artificial intelligence — AI management system. Under development.

[18] ISO/IEC 22989:2022. Artificial intelligence — Concepts and terminology.

[19] ISO/IEC 23894:2023. Information technology — Artificial intelligence — Guidance on risk management.

[20] ISO/IEC 24028:2020. Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence.
[21] ISO/IEC 24029-2:2023. Artificial intelligence — Assessment of the robustness of neural networks — Part 2: Methodology for the use of formal methods.

Institutional Publications

[22] The Future Society. "Navigating the EU AI Act: Implementation Challenges and Priorities." Policy report, 2024.
[23] Bruegel. Analysis of AI Act implementation challenges for multi-model systems. 2024–2025.
[24] MDCG 2025-6. Guidance on the relationship between the MDR/IVDR and the AI Act. Medical Device Coordination Group, June 2025.
[25] European Data Protection Board. Guidelines on automated individual decision-making and profiling for the purposes of Regulation 2016/679. WP251rev.01, as last revised February 2018.

Auburn Governance Stack

[26] Fields, R. "CRSA-1: Compositional Runtime Safety Attestation Protocol for Multi-Principal AI Systems." Auburn Governance Stack, 2026.
[27] Fields, R. "The Model Attestation Interface (MAI-1): A Normative Profile and Conformance Protocol for Foundation Model Governance." Auburn Governance Stack, 2026.
[28] Fields, R. "CTS-1: MAI-1 Conformance Test Suite." Auburn Governance Stack, 2026.
[29] Fields, R. "Auburn Governance Stack: Master Architecture Plan." Auburn Governance Stack, 2026.

Intellectual Property Declaration

The methods, logic structures, compositional risk taxonomy (CRT-1 through CRT-8), conformity assessment methodology, architecture compliance profiles, and model contract terms contained in this work are the sole property of Ryan Fields.

Public License (Non-Commercial)

- **Academic Use:** Researchers may share and use this framework for non-commercial academic purposes, provided full attribution is given to Ryan Fields.
- **No Derivatives:** No modifications or adaptations of the compositional risk taxonomy, conformity assessment methodology, or model contract terms are permitted without express written consent.

Commercial Prohibition

Commercial use of this framework is strictly prohibited. This includes, but is not limited to:

- Use within proprietary AI governance, risk management, or compliance software.
- Integration into commercial conformity assessment services or notified body assessment methodologies.
- Use by consulting firms, law firms, or advisory practices in client-facing deliverables.
- Incorporation into commercial AI system documentation or regulatory filings without license.

Ryan Fields
UncleBroFields@proton.me
fieldsryanchristopher@gmail.com

CRSA-1 EU Edition — First in EU Compliance Series
Fields, 2026