GOVERNING AGENTS

A Practitioner's Cross-Framework Reference

Mapping GDPR, the EU AI Act, NIST AI RMF, and ISO/IEC 42001 to AI Agents

Noah M. Kenney
Founder & Principal Consultant, Digital 520
President & Chief Scientist, Disruptive AI Lab
President, Ethical Tech Forum

First Edition · 2026

Copyright and Legal Disclaimer

This reference is provided for informational and educational purposes only and does not constitute legal, regulatory, compliance, or professional advice. The content reflects the author’s analysis and interpretation of agentic AI governance, law, privacy, security, and compliance frameworks as of the date of publication. It is not intended to serve as a substitute for advice from qualified legal counsel, regulatory authorities, or other licensed professionals. Artificial intelligence governance is a rapidly evolving field; laws, regulatory guidance, and enforcement practices may change over time and vary across jurisdictions.

The frameworks, control catalogs, and decision rules presented here are reference implementations and starting points for analysis. They must be adapted, validated, and approved based on the specific risk profile, industry requirements, and legal obligations of each organization. No representation or warranty is made that implementation of any approach described herein will ensure legal or regulatory compliance or prevent adverse outcomes.

Published by Digital 520. For permission requests: Info@NoahKenney.com.

About This Reference

This document is a companion reference to Kenney, N. M. (2026). Governing Intelligence: Law, Privacy, Security, and Compliance in the Age of Artificial Intelligence (1st ed.). Digital 520 (hereafter "Governing Intelligence" or "the textbook"). Governing Intelligence develops the AI Governance Stack, a five-layer operational framework comprising Data Governance, Model Governance, System Integration Governance, Control and Monitoring Governance, and Audit and Evidence Governance, as the organizing structure for responsible artificial intelligence. This reference extends that framework to a category that the textbook treats only by implication: the autonomous, tool-using AI agent.

Where Governing Intelligence reasons primarily about AI systems that produce discrete outputs in response to discrete inputs, this reference reasons about AI agents that plan, call tools, retain state across turns, delegate subtasks to other agents, and act on the world over extended horizons. The compliance consequences of that shift are substantial. A single agentic session may invoke dozens of processing operations, pull from multiple data sources, generate downstream writes to records systems, and cross lawful-basis boundaries, all within a latency budget measured in seconds. Static policy and contract structures designed for request–response systems cannot, without significant adaptation, bear that load.

This reference should be read alongside three other recent outputs from the author's research program. The first is Kenney, N. M. (2026). Runtime Enforcement of AI Governance, which argues that governance intent must be compiled into runtime controls that execute alongside the agent rather than remaining in policy documents or review gates. The second is the Operationalizing Governing Intelligence webinar (Digital 520, 2026), which presents a one-hour practitioner walkthrough of stack instantiation in a regulated enterprise. The third is Kenney, N. M. (2026). Privacy Impact Assessment of Claude Opus 4.7 (Digital 520 PIA Series), which exercises the stack against a frontier model and exposes several controls that become load-bearing only once the model is embedded in an agentic loop.

How to Use This Reference

Part I establishes why AI agents require a distinct compliance treatment and introduces an agent capability taxonomy that the rest of the reference applies consistently. Part II maps the General Data Protection Regulation (EU) 2016/679 article-by-article to agentic behaviors, identifying where traditional implementations fail and what runtime controls restore conformity. Part III overlays the Artificial Intelligence Act (Regulation (EU) 2024/1689) onto the GDPR baseline, with particular attention to the compound obligations that arise when personal data processing and high-risk use are combined. Part IV crosswalks the NIST Artificial Intelligence Risk Management Framework (AI 100-1) and its Generative AI Profile (AI 600-1). Part V crosswalks ISO/IEC 42001:2023, the first international management system standard for AI.

Part VI consolidates those four mappings into a unified crosswalk organized by agent capability. Part VII catalogs runtime enforcement patterns, policy-as-code, tool gating, memory scoping, delegation controls, and audit envelopes, that instantiate the mappings on live systems. Part VIII presents a numbered control catalog ready to be lifted into enterprise policy. Part IX records open questions, enforcement gaps, and a research agenda.

Three conventions run throughout. First, each mapping is tagged with the AI Governance Stack layer(s) it primarily instantiates, using the notation L1 through L5 as defined in Governing Intelligence chapter 1.5. Second, each GDPR, AI Act, NIST, and ISO mapping closes with an italicized "Textbook cross-reference" pointer to the chapter and section where foundational material is developed. Third, the text avoids prescriptive language where the underlying legal question is unsettled; the reference identifies the open question, surveys the authority, and leaves the judgment call with the practitioner.

A note on scope

This reference focuses on AI agents that (a) are powered by general-purpose or frontier models, (b) have tool-use capability, and (c) operate with at least session-level memory. Fully autonomous physical agents (industrial robotics, autonomous vehicles, weapon systems) inherit most of the obligations discussed here but add sector-specific safety, certification, and liability regimes that are out of scope. Domain-specific treatment is developed in Governing Intelligence chapters 17, 18, and 19.

Table of Contents

Part I. Foundations: Why Agents Change the Compliance Surface

The central claim of this reference is that AI agents are not a variant of AI systems for which existing compliance mappings can be mechanically reused. They are a distinct governance category. The regulatory texts that matter, the GDPR, the EU AI Act, the NIST AI RMF, and ISO/IEC 42001, were drafted with either traditional software or request-response AI systems in mind. Each of their core assumptions is strained, and several are broken, once a system is permitted to plan, call tools, retain state, and act over extended horizons. Governing Intelligence chapter 1 establishes the AI Governance Stack as the operational answer to the broader governance problem; this Part shows where the stack's obligations shift when the governed artifact is an agent.

1.1 What Changes When a System Becomes an Agent

The textbook defines an AI system, following the OECD and the EU AI Act, as a machine-based system that infers from inputs how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments. An AI agent satisfies that definition and adds four capabilities that together produce qualitatively different governance obligations.

First, agents plan. They decompose goals into sub-goals and select among candidate actions based on an internal representation of the task. That planning is rarely deterministic or transparent; the same prompt can produce materially different action sequences on successive runs. Second, agents use tools. They invoke external functions, APIs, code interpreters, databases, file systems, and other agents. Each invocation is a separate processing operation with its own controller or processor implications. Third, agents retain state. Session memory, working memory, long-term memory stores, and retrieval-augmented generation indexes accumulate personal data, often silently. Fourth, agents delegate. Multi-agent architectures route subtasks to specialized agents whose prompts, data access, and tool permissions may differ from the top-level agent, creating nested data flows that existing Records of Processing Activities (ROPA) cannot represent without modification.

These four capabilities transform the unit of governance. The textbook's five-layer stack still applies, but the forces on each layer change. Data Governance (L1) must accommodate data that is not merely ingested but discovered and retrieved in-flight. Model Governance (L2) must account for the fact that the model is no longer the only decision-making component; the scaffold around it often matters more. System Integration Governance (L3) becomes the load-bearing layer because tool invocations are integrations. Control and Monitoring Governance (L4) must observe events at sub-second cadence across opaque reasoning traces. Audit and Evidence Governance (L5) must record not just outputs but the planning artifacts that produced them.

Key takeaway

Agents do not relax any obligation that applies to AI systems. They extend those obligations across a larger number of processing operations per session, shift load toward integration and

monitoring controls, and introduce new obligations in the form of planning, tool-use, memory, and delegation records. The AI Governance Stack remains the correct organizing framework; the content of each layer is what changes.

1.2 A Working Definition of an AI Agent

For purposes of this reference, an AI agent is a machine-based system that (i) is powered by one or more AI models, typically including at least one general-purpose or frontier model, (ii) produces plans or action sequences in response to a goal expressed in natural or structured language, (iii) invokes external tools, APIs, or other agents to execute steps in those plans, (iv) retains at least session-scoped state and often longer-lived memory, and (v) produces effects on data, records, or the physical world beyond the return of a single generative output. Systems that satisfy (i) and (ii) but not (iii) through (v) are, for compliance purposes, AI systems rather than agents; the Governing Intelligence framework applies to them largely unmodified.

This definition is deliberately broader than some engineering usages that reserve the term “agent” for systems exhibiting long-horizon autonomy. Regulators do not currently draw that engineering distinction, and compliance obligations, in particular lawful-basis, transparency, and ADM obligations under the GDPR, are triggered by tool use and state retention well before a system becomes autonomous in any robust sense. Practitioners who adopt a narrower engineering definition risk leaving those obligations unaddressed in deployments they would not otherwise call “agentic.”

1.3 The Agent Capability Taxonomy

The rest of this reference uses a five-capability taxonomy to describe what an agent is doing at any given moment. Each capability is a separate locus of compliance obligation, and the mappings in Parts II through V tag their analyses to the capabilities they implicate.

Capability A1: Planning and Reasoning

The agent decomposes a goal into sub-goals and sequences actions. Reasoning traces (chain-of-thought, scratchpad outputs, planner internal state) may contain personal data, confidential inputs, or inferences that are themselves regulated. Planning opacity is the governance problem: explanation obligations under GDPR Article 22, transparency obligations under AI Act Article 50, and accountability obligations under NIST GOVERN-1.4 all presume that the rationale for a decision is available for inspection. Planning traces are the closest functional analog, and the textbook’s Audit and Evidence Governance layer (L5) is the natural home for their retention and disclosure rules.

Capability A2: Tool Use and External Action

The agent calls tools. Each call is (a) a processing operation under GDPR Article 4(2), (b) potentially a cross-border transfer under GDPR Chapter V if the tool is hosted outside the EEA, (c) potentially an onward disclosure subject to the controller’s contractual commitments, and (d) potentially an act performed by a processor within the meaning of Article 28. Governance of tool use is predominantly a System Integration Governance (L3) concern, extended by Control and Monitoring (L4) to enforce tool gates in real time.

Capability A3: Memory and State

The agent retains information across turns. Session memory exists for the duration of a conversation; working memory is scoped to the current task; long-term memory persists across sessions and often across users. Memory stores are processing operations, and their contents are frequently personal data. Purpose limitation (GDPR Article 5(1)(b)) and storage limitation (Article 5(1)(e)) are the principal strains; runtime enforcement of memory scoping and time-to-live (TTL) is the principal remedy. Memory sits primarily in Data Governance (L1) with dependencies on L3 for retrieval-augmented generation (RAG) pipelines.

Capability A4: Delegation and Multi-Agent Coordination

The agent routes subtasks to other agents. Sub-agents may have different system prompts, tool permissions, data access scopes, and, critically, different identities from the controller's perspective. Delegation creates nested controller-processor structures that Article 28 was not drafted to describe cleanly. Coordination is primarily an L3 and L4 concern; the ROPA obligations of Article 30 shift load to L5.

Capability A5: Adaptation and In-Context Learning

The agent adapts behavior based on retrieved context, few-shot examples, or in-context instructions. Adaptation does not alter model weights but changes behavior in ways that can cross purpose boundaries. A model that has been instructed to behave as a customer service representative is performing a different processing activity when instructed mid-session to act as a debt collector; the latter activity may require a different lawful basis, a different DPIA, and different transparency disclosures. Adaptation is primarily an L2 and L4 concern.

1.4 Mapping Agents onto the AI Governance Stack

The following summary shows how each capability loads the five layers defined in Governing Intelligence chapter 1.5. The pattern is consistent across the regulatory mappings developed in later Parts: integration and monitoring layers do the heavy lifting for agents, where they were often secondary for traditional AI systems.

Capability	L1 Data	L2 Model	L3 Integration	L4 Control & Monitoring	L5 Audit & Evidence
A1 Planning	Low	Medium	Low	High	High
A2 Tool Use	Medium	Low	High	High	Medium
A3 Memory	High	Low	Medium	High	Medium
A4 Delegation	Medium	Medium	High	High	High
A5 Adaptation	Medium	High	Medium	High	Medium

The key practitioner consequence is that organizations that have invested heavily in Data and Model Governance but treated Integration and Monitoring as checklist activities will find

themselves materially under-controlled when they deploy agents. The runtime enforcement agenda developed in Kenney (2026b) is a direct response to this shift.

1.5 Why Traditional Compliance Mappings Fail on Agents

Four structural failure modes recur across the regulatory frameworks mapped in this reference. Each is treated in detail in its chapter-specific discussion; they are introduced here to establish vocabulary.

The first is controller–processor drift. A single agent session can traverse multiple roles: the deploying organization acts as a controller for the user-facing task, becomes a processor when the agent calls a third-party API on behalf of the user, and may become a joint controller with another party when two agents coordinate. Traditional ROPA templates assume a stable controller role; agentic sessions invalidate that assumption. The textbook treats controller determination as foundational (Governing Intelligence § 8.1); this reference extends that treatment to role changes within a session.

The second is purpose-limitation decay. Agents retrieve context, receive mid-session instructions, and adapt behavior in ways that can slip across the original purpose declared in a DPIA. A chat assistant that begins by answering product questions and ends by drafting legal correspondence has changed purposes, whether or not the controller recognizes it. Runtime enforcement of purpose scopes, the mechanism developed in the runtime enforcement paper, is the only reliable remedy.

The third is transparency displacement. The GDPR and the AI Act both require meaningful information about the logic of automated decisions. Agents produce decisions through chains of reasoning, tool calls, and retrieved context that no single artifact adequately summarizes. Model cards describe the model; system cards describe the scaffold; neither describes the specific decision. A decision-level disclosure artifact, the decision envelope developed in Part VII of this reference, is required.

The fourth is evidence scarcity. High-risk AI Act systems must retain logs sufficient to reconstruct behavior. Agents generate logs at high rates; naive retention is both expensive and, under GDPR Article 5(1)(e), non-compliant. A tiered retention architecture, reasoning traces short-lived, action records longer-lived, audit envelopes longest-lived, is the pragmatic compromise.

> **Cross-reference**
> Governing Intelligence § 1.5 introduces the AI Governance Stack; § 2.1 through § 2.6 develop each layer. Runtime enforcement of the controls described in Parts II through VIII of this reference is treated comprehensively in Kenney (2026b), Runtime Enforcement of AI Governance.

Part II. The GDPR, Article-by-Article, Applied to AI Agents

This Part maps the General Data Protection Regulation (Regulation (EU) 2016/679, hereafter "GDPR") article-by-article to AI agent behaviors. For each article, the analysis identifies the agent capabilities (A1–A5) implicated, the stack layer(s) where the obligation lives (L1–L5), the ways traditional implementations strain or fail, and the runtime enforcement patterns that restore conformity. Textbook cross-references point to the foundational treatment in Governing Intelligence chapter 8.

2.1 Article 4: Definitions Applied to Agents

Article 4 carries disproportionate weight for agents because three definitions, "personal data," "processing," and "profiling", determine the scope of every subsequent obligation. Agents expand all three.

Personal data (Article 4(1)) includes any information relating to an identified or identifiable natural person. Agent memory often contains inferences, derived attributes, and behavioral fingerprints that are themselves personal data even when the raw inputs were not. A planning trace that records "user appears frustrated; suggest empathetic framing" is an inference about a natural person and is personal data. Controllers that exclude reasoning traces from their data inventories on the basis that they are "internal" misclassify them.

Processing (Article 4(2)) includes any operation performed on personal data, including retrieval, consultation, use, disclosure by transmission, and erasure. Each tool call, each retrieval from a vector store, and each message passed to a sub-agent is a processing operation. An agent session that invokes twenty tools performs at least twenty processing operations, each of which must be lawful and each of which contributes to Article 30 record-keeping obligations.

Profiling (Article 4(4)) covers any form of automated processing to evaluate, analyze, or predict aspects of a natural person. Agents routinely profile users: the planner conditions on inferred user preferences; the memory store accumulates behavioral history; the retrieval layer biases context toward past interactions. Whether or not the controller intends to profile, the technical behavior of the agent produces profiling within the meaning of Article 4(4). The Article 22 analysis in § 2.13 below follows directly.

> **Practitioner rule**
> Treat every reasoning trace as personal data until proven otherwise. The marginal cost of applying data-subject rights to traces that turn out not to contain personal data is low; the marginal cost of excluding traces that do contain personal data is a regulatory finding.

Textbook cross-reference: Governing Intelligence § 8.1 (GDPR fundamentals) and § 8.2 (DPIA triggers).

2.2 Article 5: Principles

Article 5(1) states seven principles: lawfulness, fairness, and transparency; purpose limitation; data minimization; accuracy; storage limitation; integrity and confidentiality; and accountability. Agents strain at least four of them.

Lawfulness, Fairness, and Transparency (Art. 5(1)(a))

The fairness element is strained when agents exhibit emergent behaviors inconsistent with user expectations formed at the start of a session. An agent that collects progressively more context to improve responses may be acting rationally from an engineering standpoint and unfairly from a regulatory standpoint if the user had no notice of the context accumulation. Transparency is displaced across multiple surfaces: the privacy notice, the system card, the tool-use disclosure, and decision-level explanations. No single surface currently satisfies the regulator's expectation; the decision envelope pattern in Part VII distributes the obligation across them.

Purpose Limitation (Art. 5(1)(b))

The purpose declared at collection binds subsequent processing. Agents violate the principle in two ways. First, in-session purpose drift, the assistant that migrates from product support to general legal advice without re-basing its lawful processing. Second, cross-session purpose drift, long-term memory accumulates data collected for one purpose and uses it to condition behavior in later sessions that pursue different purposes. The runtime remedy is a purpose scope attached to every data item in memory and enforced at retrieval time; retrieval that crosses scopes either blocks or triggers a compatible-purpose assessment.

Data Minimization (Art. 5(1)(c))

Agents tend toward maximization. They retrieve more context than strictly necessary to "improve" responses. Article 5(1)(c) requires adequate, relevant, and limited processing. Minimization controls include retrieval caps, context-window budgets tied to task class, and adversarial pruning of retrieved content against declared purpose.

Storage Limitation (Art. 5(1)(e))

Agent memory has a strong incentive to grow. Storage limitation requires retention periods that are no longer than necessary for the purpose. A tiered retention architecture, reasoning traces at minutes to hours, tool-call records at days to weeks, audit envelopes at years, offers a defensible compromise. Each tier needs a declared purpose, retention period, and deletion pipeline.

> **Stack layer**
> Article 5 obligations span L1 (retention and minimization), L2 (model inputs and outputs), L4 (runtime enforcement of purpose and minimization), and L5 (evidence of conformity). Accountability under Article 5(2) is specifically an L5 obligation.

Textbook cross-reference: Governing Intelligence § 8.1 and § 11.1 (Privacy by Design).

2.3 Article 6: Lawful Basis

Article 6(1) requires one of six lawful bases for every processing operation. Agents strain the requirement because a single session involves many processing operations that may require different bases. A consent-based lawful basis for the user-facing interaction does not

automatically extend to the agent's tool calls; a legitimate-interests basis for observability logging does not extend to use of those logs for model training.

The practitioner discipline is to maintain a per-operation basis map, attached to the agent at design time and checked at runtime. The map is specific enough that (a) each tool call can be associated with a basis, (b) each memory write can be associated with a basis, and (c) any processing that lacks a basis is blocked or diverted to a manual review pathway. Runtime enforcement libraries that evaluate basis at the edge of tool calls are the pattern of choice; they are developed in Kenney (2026b).

Legitimate interests (Article 6(1)(f)) deserves particular care. The balancing test required by recital 47 is difficult to run in advance for behaviors that emerge in deployment. Controllers who rely on legitimate interests for agentic processing should maintain a pre-registered balancing analysis per operation class, re-tested on a defined cadence, and should be prepared to discontinue processing where the balance tips.

Textbook cross-reference: Governing Intelligence § 8.1.

2.4 Article 7 and Article 8: Consent

Where consent is the lawful basis, it must be freely given, specific, informed, and unambiguous, and it must be as easy to withdraw as to give. Agents complicate all four elements. Specificity requires that consent be tied to particular processing purposes; agentic sessions produce processing purposes the controller may not have anticipated. Informed consent requires meaningful information about processing; agents produce processing whose logic is emergent. Withdrawal must be as easy to exercise as consent was to give; revoking consent mid-session must be effective and must propagate to downstream tool calls and memory.

A defensible consent architecture for agents separates consent to interact with the agent from consent to specific high-privacy processing activities such as retention of conversations beyond the session, use of conversations for model training, and use of profile data to adapt behavior in future sessions. Each sub-consent has its own scope, lifecycle, and withdrawal pathway. Where consent cannot realistically satisfy the specificity requirement, for example, because the set of downstream processing activities is open-ended, the controller should select a different lawful basis rather than rely on broad consent that would not withstand scrutiny.

For children (Article 8), the interaction of consent mechanics with parental authorization imposes additional runtime burdens: age-assurance before session start, re-authorization when agent behavior broadens, and special handling of any inferred-child signal discovered mid-session.

Textbook cross-reference: Governing Intelligence § 9.5 (children's privacy).

2.5 Article 9: Special Categories of Data

Article 9 prohibits the processing of special categories (racial or ethnic origin; political opinions; religious or philosophical beliefs; trade union membership; genetic data; biometric data for the purpose of uniquely identifying a natural person; data concerning health; data concerning a natural person's sex life or sexual orientation) unless one of the Article 9(2) conditions is met. Agents risk processing special categories accidentally. A planning trace that reasons about a

user's apparent religious practice, health, or sexual orientation is processing special-category data whether or not the controller intended to collect it.

Runtime controls include input filters that detect and divert special-category content, planner constraints that forbid reasoning over inferred sensitive attributes unless a basis under Article 9(2) has been established, and retrieval filters that prevent the vector store from returning documents that have been flagged as containing special categories. The explicit consent ground in Article 9(2)(a) is rarely workable for agentic processing because it requires specificity at a granularity agents cannot guarantee.

Textbook cross-reference: Governing Intelligence § 10.1 (HIPAA and health data) for the U.S. parallel.

2.6 Articles 12–14: Transparency and Information

Articles 13 and 14 require controllers to provide information about the processing at the point of collection. Agents complicate the obligation because the set of processing activities is not fully known at the point a user begins a session. The practical response is a layered transparency architecture: a short-form notice at session start; a mid-session disclosure triggered when the agent escalates capabilities (for example, when it first invokes a tool that transmits data to a third party); and a long-form reference notice that enumerates the full set of processing activities and is continuously updated.

Article 12's concision requirement is in tension with the volume of information that agents generate. The discipline is to present only the information relevant to the current interaction, with links to deeper material. The transparency log, a per-session record of which disclosures were shown and acknowledged, is necessary both to demonstrate Article 12 conformity and to support subject-access rights.

Textbook cross-reference: Governing Intelligence § 8.1 and § 8.8.

2.7 Article 15: Right of Access

Article 15 entitles a data subject to confirmation of processing, access to personal data being processed, and information about the processing's purposes, categories, recipients, retention, and logic of any automated decision-making. Agents expand the scope of "personal data being processed" beyond what controllers typically expose through subject-access portals.

An agent-aware access response must include (i) conversation transcripts, (ii) inferred attributes written to short- and long-term memory, (iii) retrieval hits drawn from the data subject's own documents or profile, (iv) planning trace excerpts where they contain personal data, and (v) tool-call records where they identify the subject. The pragmatic approach is to design the agent's storage layout so that a subject identifier can be joined across these surfaces; retrofitting join keys into an existing agent is costly.

Textbook cross-reference: Governing Intelligence § 8.1.

2.8 Article 16: Right to Rectification

Rectification in an agent context means correcting inaccurate personal data, including inferences, in memory. Two problems recur. First, inferences are not always reducible to a single field that

can be “corrected”; an inference that a user is “price-sensitive” may be distributed across multiple memory entries and model behaviors. Second, inferences produced by the model have no stable identity, rerunning the agent with corrected inputs may reproduce the error. The practitioner response is a layered rectification pipeline: surface-level overwrite of explicit memory entries, negative retrieval rules that prevent the old inference from returning to context, and audit logging of each rectification for Article 5(2) evidence.

Textbook cross-reference: Governing Intelligence § 8.1.

2.9 Article 17: Right to Erasure

Erasure is the right most strained by agent architectures. Personal data may be embedded in model-adjacent artifacts that cannot cleanly be erased: retrieval indexes (re-indexing required), fine-tuning corpora (model retraining required), reasoning traces already retained for audit (retention-limitation conflict), and inferences propagated to downstream systems (cascade erasure required).

An agent-aware erasure capability requires: (a) canonical storage of personal data with authoritative keys that can be located and removed, (b) propagation rules that identify downstream artifacts containing the data or its derivatives, (c) a model-retraining pathway for fine-tuned artifacts, and (d) an evidentiary record demonstrating that each step was performed within the statutory timeframe. Where a model was pre-trained on publicly available data containing the data subject’s information and the controller did not train the model, the controller’s erasure obligation is limited to the artifacts it controls; the controller should be transparent about this limit in the Article 13 notice.

Textbook cross-reference: Governing Intelligence § 8.1 and § 11.3.

2.10 Article 18: Right to Restriction of Processing

Restriction requires that personal data be marked so that it is stored but not actively processed. For agents, “active processing” must be interpreted to include retrieval into context. A restricted record should not appear in retrieval results, should not influence planning, and should not be visible to tool calls. Runtime enforcement of restriction flags at the retrieval and tool-call boundaries is the implementation; design-time tagging of records with a restriction-eligible flag is the prerequisite.

2.11 Article 20: Right to Data Portability

Portability applies to personal data that the data subject has provided to the controller and that is processed on the basis of consent or contract. For agents, “provided by” covers both explicit inputs and the data subject’s contributions to a session. Inferences and memories generated by the agent are outside the portability right strictly, but the controller should consider whether providing them would improve trust and reduce friction; Article 20(3) allows the controller to include more than the minimum.

Textbook cross-reference: Governing Intelligence § 8.1.

2.12 Article 21: Right to Object

Objection applies to processing based on legitimate interests or public interest, and always to direct marketing. Agents that rely on legitimate interests must honor objection, which requires runtime enforcement of per-user objection flags across every processing operation. Because agents accumulate and propagate personal data, an objection must cascade: the user's objection to processing for profile adaptation must remove profile-adaptation signals from memory, not merely stop future writes.

2.13 Article 22: Automated Decision-Making

Article 22 grants a right not to be subject to a decision based solely on automated processing, including profiling, that produces legal or similarly significant effects. The article is the single most important provision for agentic systems. Three questions dominate analysis.

First, is the decision "solely" automated? Many agent deployments include human oversight in principle but not in fact; the human reviewer approves outputs at a rate that does not permit meaningful reconsideration. Courts and data protection authorities have signaled that rubber-stamp review does not move a system out of Article 22. The practitioner discipline is to measure oversight intensity (sampling rate, override rate, time-per-decision) and to treat low-intensity oversight as equivalent to fully automated for Article 22 purposes.

Second, does the decision produce legal or similarly significant effects? Credit, insurance, employment, housing, education access, and public benefits clearly qualify. Agentic systems acting as gatekeepers in these domains, an agent that drafts a loan adverse-action notice, for example, are subject to Article 22 even if a human signs the notice. The textbook treats these domains in chapters 17 and 18.

Third, what is "meaningful information about the logic"? For agents, meaningful information must describe (a) the inputs considered, (b) the decision rule at a level a non-specialist can understand, (c) the tool calls that produced evidence relied upon, and (d) the significance and envisaged consequences. A model card is insufficient; a decision envelope that records those four elements for the specific decision is the right artifact.

> **Runtime enforcement**
> The Article 22 regime is unworkable unless the agent produces, at decision time, a structured artifact that can be inspected by the data subject and by auditors. The decision envelope pattern in Part VII is designed to meet that requirement. See also Kenney (2026b) Runtime Enforcement of AI Governance for the policy-as-code implementation.

Textbook cross-reference: Governing Intelligence § 8.3 (automated decision-making and the right to explanation).

2.14 Articles 24 and 25: Controller Responsibility and Data Protection by Design and by Default

Article 24 requires technical and organizational measures appropriate to risk. Article 25 requires data protection by design and by default. Agents shift the center of mass of appropriate measures away from pre-deployment assurance toward runtime enforcement. Design-time measures still

matter, threat modeling, DPIAs, tool-access reviews, but the variability of agentic behavior means that runtime measures must do the actual work.

Data protection by default, for agents, means (a) the smallest feasible memory by default, (b) the narrowest feasible tool-access scope by default, (c) the shortest feasible retention by default, (d) the most privacy-preserving option for any user-facing choice by default, and (e) the most restrictive option for sub-agent delegation by default. Departures from these defaults require documented justification linked to a specific purpose.

Textbook cross-reference: Governing Intelligence § 11.1 (Privacy by Design and Privacy by Default for AI Systems).

2.15 Article 28: Processors

Article 28 requires that processors act only on documented instructions from the controller, impose confidentiality on personnel, implement appropriate security measures, engage sub-processors only with controller authorization, assist with data-subject requests, assist with Article 32–36 obligations, delete or return data at end of service, and make available information necessary to demonstrate compliance. Agents stress each element.

Documented instructions: system prompts and tool schemas are the operative instructions to a processor-operated agent. They should be treated as contractual annexes, not as configuration. Sub-processor authorization: every tool endpoint that processes personal data is effectively a sub-processor; the controller must authorize the full chain. Data-subject assistance: processors must surface data held in memory, retrieval indexes, and logs in a form suitable for subject-access response. End-of-service deletion: processors must be able to erase agent-specific artifacts, memory stores, fine-tunes, retrieval indexes, on a defined schedule.

Controllers commissioning agentic processing should require processors to maintain a runtime control catalog referenced in the DPA, update it on a defined cadence, and expose it for audit. The catalog should at minimum list tool endpoints, memory surfaces, retention schedules, and the runtime enforcement controls applied to each.

Textbook cross-reference: Governing Intelligence § 16.9 (managing AI supply chains).

2.16 Article 30: Records of Processing Activities

Article 30 records must identify controllers, contact details, purposes, categories of data subjects and data, recipients, transfers, retention periods, and a general description of technical and organizational security measures. Agents explode the number of discrete processing activities to be recorded.

The pragmatic approach is a compositional ROPA. Rather than enumerate every tool call, the controller records processing classes (for example, “user query answering,” “calendar retrieval,” “third-party search”), binds each class to purposes and recipients, and maintains runtime attestation that each agent session instantiates one or more classes from the registered list. Agents that generate processing activities outside the registered classes are either blocked at runtime or escalated to human review for ROPA expansion. Runtime enforcement of ROPA coverage is a direct application of the Kenney (2026b) architecture.

Textbook cross-reference: Governing Intelligence § 8.7 (GDPR implementation in the AI Governance Stack).

2.17 Article 32: Security of Processing

Article 32 requires appropriate technical and organizational measures. For agents, the attack surface is broader than for traditional AI systems and includes prompt-injection attacks delivered via retrieved content, tool-use attacks that cause privilege escalation, memory-poisoning attacks that persist across sessions, and delegation attacks that exploit sub-agent trust boundaries. The textbook develops the AI threat landscape in chapter 13; agent-specific threats inherit those patterns and add tool-call and memory-specific ones.

The minimum-viable Article 32 posture for agents includes: defense-in-depth input filtering (separating system instructions from user content and retrieved content), tool gating with allowlists per task class, memory isolation between users, rate limiting at every tool boundary, continuous adversarial evaluation, and incident response procedures keyed to agent-specific indicators of compromise.

Textbook cross-reference: Governing Intelligence chapters 12, 13, and 14.

2.18 Articles 33 and 34: Breach Notification

Breach notification obligations attach to "personal data breaches," defined as breaches of security leading to accidental or unlawful destruction, loss, alteration, unauthorized disclosure of, or access to, personal data. Agent-specific breach types include successful prompt-injection causing data exfiltration, memory-scope violations (one user's data returned to another), tool misuse exposing data to unauthorized sinks, and delegation leaks in multi-agent architectures. Practitioners should pre-script notification templates for each class to meet the 72-hour deadline under Article 33(1).

Textbook cross-reference: Governing Intelligence § 12.8 (incident response for AI systems).

2.19 Article 35: Data Protection Impact Assessment

Article 35 requires a DPIA where processing is likely to result in a high risk to the rights and freedoms of natural persons. The WP29 guidelines identify nine criteria; agentic processing frequently hits at least five (evaluation or scoring, automated decision-making with significant effect, systematic monitoring, data processed on a large scale, data concerning vulnerable subjects).

An agent-aware DPIA differs from the textbook's baseline (Governing Intelligence § 8.2) in four specific ways. First, it enumerates the agent's tool roster and treats each tool as a distinct processing activity. Second, it describes the agent's memory architecture and retention rules. Third, it describes the delegation graph and the trust boundaries between agents. Fourth, it describes the runtime enforcement architecture that instantiates the DPIA's safeguards in production. Without the fourth element, the DPIA remains a paper exercise; with it, the DPIA is load-bearing.

The author's Privacy Impact Assessment of Claude Opus 4.7 (Kenney, 2026c) demonstrates the extended DPIA template on a frontier model deployed in agentic contexts; practitioners can use that assessment as a reference instantiation.

Textbook cross-reference: Governing Intelligence § 8.2 and § 11.5.

2.20 Articles 44–49: International Transfers

Transfers of personal data to third countries are permitted only under an adequacy decision, appropriate safeguards, or a derogation. Agents present transfer challenges that static architectures did not. Tool calls may route data to endpoints outside the EEA at runtime and in response to model decisions; retrieval may pull context from indexes that mirror to non-EEA regions; delegation to sub-agents may invoke model endpoints in third countries without the user or controller observing the fact.

The practical architecture is a transfer-aware tool registry that tags every tool endpoint with its hosting jurisdiction and the applicable transfer mechanism (adequacy, SCC, BCR, derogation). Runtime enforcement blocks tool calls that would produce an unauthorized transfer and substitutes compliant alternatives where possible. The Schrems II supplementary-measures analysis (Governing Intelligence § 8.4) should be run at the tool level, not only at the service level.

Textbook cross-reference: Governing Intelligence § 8.4.

2.21 Article 82: Liability and the Right to Compensation

Article 82(2) distinguishes controller liability ("any infringement") from processor liability (processor-specific obligations or acting outside controller instructions). In agentic settings, the distinction is blurred because the agent can produce effects neither controller nor processor instructed. The contractual regime, indemnification clauses, liability caps, insurance, should be drafted to reflect agentic failure modes: prompt-injection exfiltration, memory-scope violations, delegation leaks, and tool-misuse events. Controllers should consider requiring processor liability for failure to implement runtime controls described in the DPA.

2.22 Summary: GDPR × Agent Capability

The following matrix summarizes which agent capabilities predominantly implicate each GDPR provision treated above. It should be read as a prioritization aid rather than an exhaustive mapping.

GDPR Provision	A1 Planning	A2 Tool Use	A3 Memory	A4 Delegation	A5 Adaptation
Art. 4 Definitions	●	●	●	●	●
Art. 5 Principles	●	●	●	●	●
Art. 6 Lawful Basis		●	●	●	●
Art. 7 Consent		●	●		●

GDPR Provision	A1 Planning	A2 Tool Use	A3 Memory	A4 Delegation	A5 Adaptation
Art. 9 Special Cat.	●	●	●		●
Art. 13–14 Transparency	●	●		●	
Art. 15 Access	●	●	●	●	
Art. 17 Erasure			●	●	●
Art. 22 ADM	●	●		●	●
Art. 25 DPbDD	●	●	●	●	●
Art. 28 Processor		●	●	●	
Art. 30 ROPA	●	●	●	●	
Art. 32 Security	●	●	●	●	●
Art. 35 DPIA	●	●	●	●	●
Art. 44–49 Transfers		●	●	●	

Part III. The EU AI Act Overlaid on the GDPR Baseline

The Artificial Intelligence Act (Regulation (EU) 2024/1689, hereafter "AI Act") governs AI system providers, deployers, importers, and distributors. It is not a privacy regulation; the GDPR continues to apply to any personal data processing performed by or via an AI system. For agents that process personal data, the AI Act obligations combine with GDPR obligations to produce a compound regime that is tighter than either regulation in isolation. This Part maps the AI Act onto agents with that compound regime in mind.

The reader is assumed to be familiar with the legislative framing developed in Governing Intelligence chapter 4, including the risk-based classification (prohibited, high-risk, limited-risk, minimal-risk), the obligations applicable to providers of general-purpose AI (GPAI) models, and the compliance timeline. This Part focuses on where agentic behavior changes the analysis.

3.1 Article 5: Prohibited Practices and Agentic Use

Article 5 prohibits specific AI practices, including manipulative or exploitative systems that materially distort behavior, social scoring by public or private actors, untargeted facial-recognition scraping, emotion recognition in workplaces and educational institutions (subject to medical or safety exceptions), biometric categorization based on sensitive attributes, and real-time remote biometric identification in publicly accessible spaces for law enforcement (subject to narrow exceptions). Agents raise risk of prohibited practices in two ways.

First, agents increase the plausibility and scale of manipulative interaction. An agent that adapts its persuasive strategy to observed user traits, and that does so at scale, can fall within Article 5(1)(a) even where individual interactions do not. Second, agents can combine otherwise-permitted capabilities into prohibited composites; an agent that uses emotion inference and employment context to prioritize calendar holds for termination meetings, for example, implicates Article 5(1)(f) even though emotion inference and calendar access are separately lawful.

The practitioner discipline is to run an Article 5 screen at agent design and at every material change, new tool, new sub-agent, new domain. Where a screen raises concerns, the affected capability should be removed or the agent should be excluded from the Union market.

Textbook cross-reference: Governing Intelligence § 4.3.

3.2 High-Risk Classification for Agents (Articles 6 and 7, Annex III)

Article 6 classifies a system as high-risk either because it is a safety component of a product covered by specified Union harmonization legislation (Article 6(1)) or because it is listed in Annex III (Article 6(2)). Annex III covers biometrics, critical infrastructure, education, employment, access to essential services, law enforcement, migration, and administration of justice and democratic processes.

Agents routinely operate in Annex III domains without being recognized as "Annex III systems." An agentic assistant embedded in a human-resources platform is deployed in the employment domain (Annex III(4)) for the purpose of filtering or evaluating candidates, recommending task allocation, or monitoring performance. The agent's provider may argue that the agent is a general-

purpose assistant, not an employment-specific system; deployers should resist that framing when the actual use qualifies.

Article 6(3), added in the final trilogue, excludes systems that perform narrow procedural tasks, improve the result of previously completed human activity, detect decision-making patterns without substituting for the decision, or perform preparatory tasks. Agents can attempt to qualify under 6(3), but the exception is narrow and does not apply where the system performs profiling (Article 6(3), final paragraph). Given that agents routinely profile within the meaning of GDPR Article 4(4), the 6(3) exception is rarely available for agents.

Where an agent is high-risk, the full suite of Chapter III obligations applies: risk management (Art. 9), data governance (Art. 10), technical documentation (Art. 11), record-keeping (Art. 12), transparency to deployers (Art. 13), human oversight (Art. 14), accuracy, robustness, and cybersecurity (Art. 15), a quality management system (Art. 17), post-market monitoring (Art. 72), and serious-incident reporting (Art. 73). Governing Intelligence § 4.4 enumerates these obligations; the remainder of this section identifies where agentic behavior creates additional implementation work.

Article 9: Risk Management and the Agent Lifecycle

Article 9 requires a continuous, iterative risk management system over the entire lifecycle. For agents, “over the lifecycle” must include the operational phase in which the agent encounters adversarial users, new tool integrations, and emergent capabilities. The risk management artifact is not a one-time document; it is a living object maintained by an operational team, refreshed on the cadence of material changes (tool additions, policy updates, model swaps), and anchored to the DPIA (GDPR Art. 35) through shared risk taxonomy.

Article 10: Data Governance and Quality

Training, validation, and testing datasets must be relevant, sufficiently representative, and to the best extent possible free of errors and complete. For agents, the data surface includes not only model training data but also retrieval corpora, memory stores, and the tool outputs that constitute de facto training signal through in-context learning. Article 10 should be read to extend to that broader surface when the surface materially influences agent behavior. Retrieval corpus curation and memory hygiene become Article 10 obligations, not merely good practice.

Article 11 and Annex IV: Technical Documentation

Annex IV enumerates the required content of technical documentation. Agents require extensions of each element. The “general description of the system” should enumerate tools, memory surfaces, and sub-agents. The description of the “elements and processes for its development” should include the scaffolding code, prompt templates, and policy-as-code modules. The description of monitoring and control systems should explicitly describe runtime enforcement. The metrics describing accuracy and robustness should include agentic metrics (tool-call precision, memory-scope conformance, plan-adherence rate).

Article 12: Record-Keeping and Logs

Article 12 requires automatic recording of events (logs) over the lifetime. For agents, event granularity is both a conformity obligation and a privacy hazard. The tiered retention architecture

introduced in § 2.22, reasoning traces shortest, tool-call records medium, audit envelopes longest, provides a defensible compromise between Article 12 and GDPR Article 5(1)(e).

Article 13: Transparency to Deployers

Providers must supply deployers with instructions for use sufficient to enable proper use. For agents, instructions must describe the tool roster, the memory architecture, the delegation graph, the policy-as-code controls shipped with the system, and the metrics deployers should monitor to maintain conformity. Providers should publish the instructions for use as a living document updated on material changes rather than as a static PDF.

Article 14: Human Oversight

Human oversight must be effective. For agents, effective oversight is not possible at the granularity of individual tool calls; oversight must operate at the granularity of plans and decisions. The oversight architecture should include a planner-level review gate for plans that cross risk thresholds, a decision-level review gate for decisions within Article 22 scope, and an aggregate-level review gate that samples sessions for audit. Automation bias is the principal Article 14 failure mode for agents; oversight procedures must include explicit mitigations (dissent surfaces, counterfactual prompts, calibration of reviewers over time).

Article 15: Accuracy, Robustness, and Cybersecurity

Agents introduce robustness challenges beyond traditional AI: prompt-injection, tool-chain attacks, and adversarial retrieval. Article 15 should be read to require resilience against these vectors as a precondition of conformity. The red-teaming program developed in Governing Intelligence § 13.8 and § 14.9 should be expanded to agent-specific attack catalogs (OWASP LLM Top 10 provides a useful baseline).

3.3 Article 50: Transparency Obligations

Article 50 imposes transparency obligations on certain AI systems regardless of risk class. Systems that interact with natural persons must disclose that fact; systems that generate synthetic content must mark it as AI-generated; emotion recognition and biometric categorization systems must inform data subjects; and deep-fake content must be disclosed. Agents implicate each of these. An agentic customer assistant must disclose its nature at the start of interaction; an agent that writes marketing copy must mark output as AI-generated (subject to the artistic-work exception); an agent that adapts responses based on inferred emotional state must disclose that inference.

Interaction with GDPR Articles 13 and 14 is important. A single disclosure can satisfy both regimes if drafted to include the information each requires; a disclosure that satisfies only one is incomplete.

Textbook cross-reference: Governing Intelligence § 4.4.

3.4 Articles 51–55: General-Purpose AI Models Powering Agents

Articles 51 through 55 impose obligations on providers of general-purpose AI (GPAI) models, with additional obligations for models presenting systemic risk (those exceeding the FLOP threshold

or otherwise designated). Agents are frequently powered by GPAI models. Three compound effects matter for agent builders.

First, the GPAI provider's technical documentation must include information about training data (Article 53(1)(d)) and a public summary (Article 53(1)(d), second subparagraph). Agent builders should review that documentation and retain it as evidence in their own technical file. Second, downstream providers that integrate a GPAI model into their agentic system may themselves become providers of a derived high-risk system; integration work does not shield them from provider obligations. Third, where the GPAI model presents systemic risk, the provider's obligations under Article 55, state-of-the-art model evaluation, systemic risk assessment, cybersecurity, extend to the capability profile that ships inside the agent.

Textbook cross-reference: Governing Intelligence § 4.5.

3.5 Compound Obligations: GDPR × AI Act

Where personal data processing is combined with high-risk AI Act use, obligations compound. The following table captures the most important compound effects for agentic systems.

Interaction	GDPR obligation	AI Act obligation	Compound effect
DPIA × Risk Management	Art. 35 DPIA prior to high-risk processing	Art. 9 continuous risk management	DPIA becomes a living artifact; risk-management cycle drives re-DPIA triggers
Transparency × Disclosure	Arts. 12–14 information	Art. 50 AI-specific disclosure	Unified disclosure architecture; single user-facing notice satisfying both regimes
Logging × Storage Limitation	Art. 5(1)(e) storage limitation	Art. 12 event logging	Tiered retention architecture; per-tier retention and deletion evidence
ADM × Human Oversight	Art. 22 ADM protections	Art. 14 human oversight	Decision-envelope artifact serving both; oversight intensity metrics
Transfers × Market Access	Chapter V transfer mechanisms	AI Act conformity presumption	Transfer-aware tool registry; jurisdictional fallbacks

Compound regime principle

Neither the GDPR nor the AI Act is sufficient in isolation for agentic systems that process personal data in high-risk domains. Design documentation should be built to satisfy both in a

single artifact set; design-time decisions that satisfy one regime while ignoring the other will fail audit under the neglected regime.

Part IV. NIST AI Risk Management Framework Applied to Agents

The NIST Artificial Intelligence Risk Management Framework (AI 100-1, hereafter "AI RMF") is a voluntary, consensus-based framework organized around four core functions: Govern, Map, Measure, and Manage. The framework is complemented by the Generative AI Profile (AI 600-1), which extends the core to generative systems. Governing Intelligence § 5.5 and § 14.1 develop the framework's overall structure; this Part maps its subcategories to agent-specific risks and controls.

4.1 GOVERN: Culture, Process, and Accountability

GOVERN establishes the organizational context for responsible AI. Its subcategories most relevant to agents are GOVERN-1.1 (legal and regulatory requirements understood and documented), GOVERN-1.4 (processes for human oversight), GOVERN-2.1 (roles and responsibilities documented), GOVERN-3.2 (decision-making policies documented), GOVERN-4.1 (organizational practices for risk management), GOVERN-5.1 (policies for third-party risk), and GOVERN-6.1 (policies for incident response).

Agent-specific instantiations: GOVERN-1.1 obligations include documenting GDPR, AI Act, sector-specific, and cross-border requirements that apply to the agent's tool roster; GOVERN-1.4 requires specifying human-oversight intensity and procedures per decision class; GOVERN-5.1 requires a full sub-processor chain including every tool endpoint and sub-agent; GOVERN-6.1 requires agent-specific IR playbooks for prompt-injection, memory-scope violation, tool-misuse, and delegation-leak events.

4.2 MAP: Context, Use Case, and Impact Analysis

MAP establishes the context for the specific AI system. MAP-1.1 (intended purposes and settings), MAP-2.2 (categories of stakeholders), MAP-3.1 (potential benefits and costs), MAP-3.4 (identification of known and potential harms), and MAP-4.1 (third-party systems and their potential impact) are the subcategories that do the most work for agents.

MAP-1.1 for agents must describe the open-endedness of intended purposes, agents produce capabilities the provider did not fully anticipate. The Map artifact should therefore include both the declared purposes and the mechanisms by which purposes are constrained at runtime. MAP-3.4 for agents should enumerate agent-specific harms: privacy breach via memory scope violation, financial or legal harm via tool misuse, reputational harm via deepfake generation, and autonomy harms via manipulative adaptation.

4.3 MEASURE: Quantitative and Qualitative Assessment

MEASURE evaluates the identified risks. MEASURE-1.1 (metrics identified), MEASURE-2.2 (evaluation involves diverse perspectives), MEASURE-2.3 (test sets representative), MEASURE-2.6 (system tested in deployment context), MEASURE-3.1 (mechanisms to track identified risks), and MEASURE-4.2 (measurement results evaluated by independent experts) are the key subcategories.

Agent-specific measurements include: tool-call precision (does the agent invoke the right tool for the task?); memory-scope conformance (does the agent retrieve only what it is authorized to retrieve?); plan-adherence rate (does the agent execute the plan it committed to?); refusal quality (does the agent refuse appropriately and not over-refuse?); and post-decision explanation quality (are decision envelopes complete and accurate?). Each metric should have a baseline, a target, and a trigger threshold for re-evaluation.

4.4 MANAGE: Risk Treatment and Continuous Improvement

MANAGE handles the risks. MANAGE-1.1 (risks prioritized), MANAGE-1.3 (responses planned), MANAGE-2.2 (monitoring mechanisms), MANAGE-4.1 (post-deployment monitoring), and MANAGE-4.3 (incident response processes) are the key subcategories.

Agent-specific risk treatment includes the runtime enforcement catalog from Part VII of this reference, policy-as-code modules that instantiate declared risk responses at the tool and memory boundaries. Post-deployment monitoring must run continuously; a quarterly review cycle that is adequate for traditional systems is inadequate for agents that can acquire new capabilities overnight via a tool-roster change or a model update.

4.5 The Generative AI Profile (AI 600-1) for Agent-Powering Models

The Generative AI Profile enumerates risks specific to generative systems: confabulation, dangerous or violent recommendations, data privacy, environmental impact, harmful bias, human-AI configuration, information integrity, information security, intellectual property, obscene or CSAM content, toxicity or harassment, and value chain and component integration. Each risk applies to agents and is amplified by agentic capabilities.

Confabulation in particular takes on new weight. A traditional generative system that confabulates produces an incorrect answer; an agent that confabulates produces an incorrect action. The consequence class shifts from information harm to decision harm. Runtime enforcement that intercepts high-confidence-confabulation indicators, for example, tool calls whose inputs cannot be grounded in retrieved context, is the pragmatic mitigation. Information security risks similarly amplify; a tool with write access to a records system converts a prompt-injection attack from an information hazard into a persistence hazard.

4.6 NIST AI RMF × AI Governance Stack

AI RMF Function	Dominant Stack Layer(s)	Agent-specific emphasis
GOVERN	L5 (Audit & Evidence), L1 (Data)	Policies for tool-roster change, memory retention tiering, sub-processor chain including tool endpoints
MAP	L1 (Data), L2 (Model), L3 (Integration)	Open-ended purpose enumeration; agent-specific harm catalog; tool and sub-agent registry

AI RMF Function	Dominant Stack Layer(s)	Agent-specific emphasis
MEASURE	L4 (Control & Monitoring), L5 (Audit)	Agentic metrics (tool-call precision, memory-scope conformance, plan-adherence rate)
MANAGE	L3, L4, L5	Runtime enforcement of declared responses; continuous monitoring; agent-specific IR playbooks

Textbook cross-reference: Governing Intelligence § 5.5 (NIST standards) and § 14.1 (AI RMF overview).

Part V. ISO/IEC 42001 Applied to Agents

ISO/IEC 42001:2023 is the first international management system standard for artificial intelligence. It follows the harmonized high-level structure (HLS) common to ISO management system standards (ISO 9001, 27001, 27701), adding AI-specific controls in Annex A. The standard is certifiable; certification is rapidly becoming a procurement expectation in regulated industries. Governing Intelligence § 14.3 and § 16 develop the standard's role in enterprise governance; this Part identifies where agents change the implementation.

5.1 Clauses 4–10: The Management System Structure

The HLS clauses require context of the organization (4), leadership (5), planning (6), support (7), operation (8), performance evaluation (9), and improvement (10). Agents impose specific substantive content at each clause. In Clause 4, the interested parties include not only end-users and regulators but tool providers and sub-agent operators. In Clause 6, the AI objectives must include agent-specific objectives (tool-call precision, decision-envelope coverage). In Clause 8, operational planning must include a change-control regime that treats tool-roster changes, model swaps, and prompt updates as planned changes with risk assessment, not as configuration.

Clause 9 (performance evaluation) interacts particularly strongly with runtime enforcement. The standard's monitoring, measurement, analysis, and evaluation obligations can be partially satisfied by runtime telemetry; the policy-as-code telemetry described in Kenney (2026b) produces evidence suitable for Clause 9.1 review. Clause 10 (improvement) requires corrective action; for agents, corrective action often takes the form of tightening runtime controls rather than retraining models.

5.2 Annex A: AI-Specific Controls and Agentic Implementation

Annex A controls cluster around AI policy and objectives (A.2), internal organization (A.3), resources (A.4), impact assessment (A.5), AI system life cycle (A.6), data for AI (A.7), information for interested parties (A.8), use of AI systems (A.9), and third-party relationships (A.10). The following control-by-control extensions matter most for agents.

A.5.4 (assessment of AI system impact) should be integrated with GDPR Art. 35 DPIAs and AI Act Art. 9 risk management to avoid three parallel artifacts. A.6 controls (life cycle) must be adapted to accommodate tool additions as material life-cycle events. A.7.3 (data quality) extends to retrieval corpora and memory stores. A.8.2 (external reporting) combines with AI Act Art. 50 to produce a single external-disclosure architecture. A.10.2 (responsibilities within the AI system life cycle) must assign ownership for tool endpoints, sub-agents, and runtime control modules.

5.3 ISO/IEC 42001 × AI Governance Stack

ISO/IEC 42001 Clause/Annex	Dominant Stack Layer(s)	Agent-specific emphasis
Clause 6 Planning	L2, L5	Agent-specific objectives; change control for tools and sub-agents

ISO/IEC 42001 Clause/Annex	Dominant Stack Layer(s)	Agent-specific emphasis
Clause 8 Operation	L3, L4	Operational planning includes runtime enforcement lifecycle
Clause 9 Performance Evaluation	L4, L5	Runtime telemetry produces Clause 9 evidence
Annex A.5 Impact Assessment	L1, L5	Integrated DPIA/Art. 9/A.5 artifact
Annex A.6 Life Cycle	All layers	Tool additions and model swaps as life-cycle events
Annex A.7 Data for AI	L1	Extends to retrieval corpora and memory
Annex A.8 Information for Interested Parties	L5	Unified disclosure architecture with AI Act Art. 50
Annex A.10 Third-Party Relationships	L3, L5	Sub-processor chain including tool endpoints and sub-agents

Textbook cross-reference: Governing Intelligence § 14.3 and chapter 16.

Part VI. Unified Crosswalk: Agent Capability × Cross-Framework Obligations

This Part consolidates the mappings developed in Parts II through V into a single unified crosswalk indexed by agent capability. The crosswalk is not a substitute for the per-article analysis; it is a navigational aid to help practitioners move from an observed agent behavior to the applicable obligations across all four frameworks.

6.1 The Master Crosswalk

Rows represent the agent capability taxonomy introduced in § 1.3. Columns map to the four frameworks. Cell contents identify the provisions and control families most directly engaged, using a compact notation.

Capability	GDPR	EU AI Act	NIST AI RMF	ISO/IEC 42001
A1 Planning & Reasoning	Art. 5(1)(a), Art. 22, Art. 15	Art. 14 oversight, Art. 12 logs, Annex IV docs	GOVERN-1.4, MEASURE-2.6, MANAGE-4.1	Clause 7.4, A.6.2.3, A.8.2
A2 Tool Use & External Action	Art. 6, Art. 28, Art. 30, Art. 32, Arts. 44–49	Art. 9, Art. 13, Art. 15, Art. 50	GOVERN-5.1, MAP-4.1, MANAGE-2.2	A.10.2, A.10.3, A.6.2.2, Clause 8.2
A3 Memory & State	Arts. 5(1)(b)(c)(e), Art. 6, Art. 17, Art. 18	Art. 10, Art. 12, Art. 15	MAP-3.4, MEASURE-3.1, MANAGE-4.1	A.7.2, A.7.4, Clause 8.1
A4 Delegation & Multi-Agent	Art. 4, Art. 28, Art. 30, Art. 32	Art. 3(3) provider chain, Art. 25, Annex IV	GOVERN-2.1, GOVERN-5.1, MAP-4.1	A.10 series, A.3.2
A5 Adaptation & In-Context Learning	Arts. 5(1)(a)(b), Art. 13, Art. 22	Arts. 5, 9, 14, 50	MAP-3.4, MEASURE-2.6, MANAGE-1.3	A.6.2, A.7, A.8

6.2 Stack-Layer Aggregation

The following summary aggregates the cross-framework obligations to the AI Governance Stack layers, supporting organizational design decisions about where to place control ownership.

Stack Layer	Primary GDPR	Primary AI Act	Primary AI RMF	Primary ISO 42001
L1 Data Governance	Arts. 5, 9, 17	Art. 10	MAP, MEASURE	A.7, A.5

Stack Layer	Primary GDPR	Primary AI Act	Primary AI RMF	Primary ISO 42001
L2 Model Governance	Art. 25	Arts. 11, 15	MAP, MANAGE	A.6, Clause 8.3
L3 System Integration	Arts. 28, 30, 32, 44–49	Art. 13, Annex IV	GOVERN-5.1, MAP-4.1	A.10, Clause 8
L4 Control & Monitoring	Arts. 5(2), 25, 32	Arts. 14, 15, 72	MEASURE, MANAGE	Clauses 9, 10
L5 Audit & Evidence	Arts. 5(2), 24, 30, 33–34	Arts. 12, 17, 73	GOVERN, MEASURE-4.2	Clauses 9.2, 9.3, A.8

The recurring pattern is that L3 and L4 bear the majority of agent-specific load across all four frameworks. Organizations whose current governance investment concentrates on L1 and L2 should expect the most work at the integration and monitoring layers to bring existing AI controls up to agent-readiness.

Part VII. Runtime Enforcement Patterns for Agent Governance

The mappings in Parts II through VI describe what the regulatory regimes require; this Part describes how to instantiate those requirements in code and operations at runtime. Each pattern is a reference implementation, not a product specification; adoption requires adaptation to the specific stack, threat model, and regulatory posture of the deploying organization. The patterns collectively elaborate the architecture introduced in Kenney (2026b), Runtime Enforcement of AI Governance.

7.1 Policy-as-Code for Agents

Policy-as-code expresses governance rules as executable predicates that run alongside the agent and either approve, deny, or modify proposed actions. The pattern separates policy (what is allowed) from mechanism (how the agent operates), which allows policy to be updated without redeploying the agent and allows policy to be audited independently.

For agents, policy-as-code predicates evaluate at four boundaries: the intake boundary (user inputs), the planning boundary (proposed plans), the tool boundary (proposed tool calls with arguments), and the output boundary (proposed outputs). At each boundary, a policy engine receives a structured description of the proposed action and returns a decision. The engine's configuration, the rule set, is the compiled expression of the organization's declared obligations.

Representative rule classes include purpose-scope rules (block operations outside the declared purpose set), jurisdictional rules (block tool calls that would produce unauthorized transfers), special-category rules (divert operations that appear to process Art. 9 data without a basis), objection rules (enforce Art. 21 objections across operations), and decision-class rules (require a decision envelope for operations within Art. 22 scope).

7.2 Tool Gating

Tool gating restricts which tools the agent can invoke in a given context. Gating operates at three scopes: per-tenant (the deployer's configuration), per-user (user entitlements and objections), and per-task (policy-determined capability restrictions for the current goal). Gating is implemented as an allowlist checked at each tool call. Denied calls either fail closed or are routed to a human reviewer, depending on policy class.

Tool registries should record, for each tool: purpose categories supported, data categories ingressed and egressed, hosting jurisdiction and transfer mechanism, rate limits and quotas, expected failure modes and recovery, and the sub-processor chain reached through the tool. The registry is both an operational artifact (used by the gate) and an evidence artifact (used in ROPA and DPIA materials).

7.3 Memory Scoping and Retention

Memory scoping attaches a scope identifier to every memory entry and a scope predicate to every retrieval. Scopes include session, user, tenant, and purpose. A retrieval that would return content across an incompatible scope either fails or triggers a compatibility assessment under GDPR Art.

6(4). Retention is implemented as TTL per scope; deletion pipelines emit evidence to the audit layer.

The tiered retention architecture repeatedly referenced in this document, reasoning traces shortest, tool-call records medium, audit envelopes longest, is a specific instantiation of scope-based retention. Each tier is a scope with its own TTL and its own purposes. Consolidated evidence that each tier's TTL is enforced satisfies both GDPR Art. 5(1)(e) and ISO/IEC 42001 Clause 9.1.

7.4 Delegation Controls

Delegation controls govern which sub-agents a top-level agent may invoke, with what data, for what purposes, and under what oversight. The control pattern mirrors tool gating but adds identity and attestation requirements: each sub-agent presents a signed capability profile, the top-level agent verifies the profile against the current task, and the runtime records the delegation chain in the audit envelope.

Delegation controls also propagate user consent and objection flags down the chain. A user who objects to profiling must not see the objection erased because a sub-agent, unaware of the flag, performs an operation the top-level agent would have refused. Propagation is implemented by passing a structured principal context on every sub-agent invocation.

7.5 Audit Envelopes and Decision Envelopes

An audit envelope is a structured record of an agent session sufficient to reconstruct the session's processing activities, tool calls, memory operations, and decisions. A decision envelope is a structured record of a specific decision sufficient to support Art. 22 explanation, Art. 50 AI Act disclosure, and Art. 15 access responses. Envelopes are generated at decision time, signed, and retained according to the tier rules in § 7.3.

The minimum content of a decision envelope includes: the decision identifier and timestamp; the data subject identifier (where determinable); the processing purpose; the lawful basis; the inputs and retrieved context (or hashes thereof); the planner-level rationale at a human-readable granularity; the tool calls and their outputs; the output and any significance indicator (for example, "adverse decision"); and the human-oversight record, if any.

7.6 Continuous Evaluation and Drift Detection

Agents drift. A stable model embedded in an evolving tool roster produces different behavior over time even when weights are unchanged. Continuous evaluation pipelines run agent regression suites on a defined cadence, compare outputs to baselines, and raise alerts when drift exceeds thresholds. The suites should include red-team cases that exercise adversarial vectors identified in Governing Intelligence chapters 13 and 14, as well as fairness and privacy cases. Drift that exceeds a threshold triggers a re-evaluation gate: re-DPIA, re-risk-assessment, re-certification as applicable.

7.7 Incident Response Integration

Incident response for agents must be keyed to agent-specific indicators. The playbook inherits the generic structure from Governing Intelligence § 12.8 and adds four agent-specific triggers: prompt-injection detection exceeding baseline, memory-scope violations detected by the policy engine, tool-misuse events detected at the tool boundary, and delegation-chain anomalies detected by audit-envelope inspection. Each trigger has defined notification, containment, eradication, recovery, and post-incident activities. Sub-72-hour GDPR notification is feasible only if the triggers are pre-wired to notification templates.

The governance contract

A deployed agent is a governance contract enacted continuously. Runtime enforcement is the execution mechanism. Policy documents, DPIAs, risk assessments, and management-system artifacts are the specification; without an enforcement substrate, the specification is aspirational. Every control catalog in Part VIII assumes an enforcement substrate is in place.

Part VIII. Agent Control Catalog

This Part presents a numbered catalog of runtime-enforceable agent controls organized by AI Governance Stack layer. The catalog is directly compatible with the enterprise compliance program architecture developed in Governing Intelligence § 16.2 and with the PIA template exercised in Kenney (2026c). Controls are written as outcomes; specific implementations vary by platform.

8.1 Layer 1 Controls, Data Governance

AG-L1-01. Every item written to short-term, session, or long-term memory is tagged with a scope identifier, a purpose identifier, and a lawful basis identifier.

AG-L1-02. Every retrieval operation evaluates scope and purpose compatibility before returning results; incompatible retrievals fail closed.

AG-L1-03. Memory stores implement tiered retention with per-tier TTL, automated deletion pipelines, and deletion-evidence emission to the audit layer.

AG-L1-04. Retrieval corpora and vector indexes are subject to the same data-quality, lineage, and erasure controls as training data.

AG-L1-05. Personal data in memory is addressable by canonical subject identifier sufficient to support Arts. 15, 16, 17, 18, and 20 requests across all memory surfaces.

AG-L1-06. Special-category data detected in user inputs or tool outputs is either filtered before memory write or tagged with a basis under Art. 9(2) and elevated handling rules.

AG-L1-07. Inferences written to memory are marked as inferences, attributed to the inferring component, and subject to Art. 16 rectification including negative retrieval rules.

8.2 Layer 2 Controls, Model Governance

AG-L2-01. Every model invoked by the agent is registered with purpose, data-category ingress and egress, training data provenance (or a summary per AI Act Art. 53), and applicable model-risk classifications.

AG-L2-02. Fine-tuning corpora used for agent behaviors are inventoried and subject to GDPR erasure obligations with a documented retraining or equivalent pathway.

AG-L2-03. Model swaps are material changes requiring re-risk-assessment, re-DPIA where applicable, and re-evaluation against the agentic metric set in § 4.3.

AG-L2-04. System prompts and other steering artifacts are versioned, reviewed, and treated as Art. 28(3)(a) documented instructions where a processor relationship is in effect.

8.3 Layer 3 Controls, System Integration Governance

AG-L3-01. A tool registry enumerates every tool available to the agent with its purpose categories, data categories, hosting jurisdiction, transfer mechanism, rate limits, failure modes, and sub-processor chain.

AG-L3-02. Tool gating enforces per-tenant, per-user, and per-task allowlists at each tool call; denied calls fail closed or route to human review.

AG-L3-03. Tool calls that would produce a transfer of personal data outside the EEA are blocked unless a valid transfer mechanism (adequacy, SCC, BCR, or derogation) is attested in the tool registry.

AG-L3-04. Sub-agent delegation is subject to identity and capability attestation; delegation chains are recorded in the audit envelope.

AG-L3-05. Retrieval-augmented generation pipelines isolate user content, retrieved content, and system instructions to reduce prompt-injection risk per § 2.17.

AG-L3-06. Tool schemas are controlled as technical documentation under AI Act Annex IV where the agent is high-risk.

8.4 Layer 4 Controls, Control and Monitoring Governance

AG-L4-01. A policy-as-code engine evaluates every proposed action at the intake, planning, tool, and output boundaries; evaluation outcomes are emitted to the audit layer.

AG-L4-02. Objection and consent-withdrawal signals are propagated to all active sessions for the subject, and to downstream processing artifacts (memory, retrieval indexes) according to documented cascade rules.

AG-L4-03. Agentic metrics (tool-call precision, memory-scope conformance, plan-adherence rate, refusal quality, decision-envelope coverage) are measured continuously with baselines, targets, and trigger thresholds.

AG-L4-04. Drift detection monitors deviations from baseline behavior and triggers re-evaluation gates when thresholds are exceeded.

AG-L4-05. Human oversight intensity is measured (sampling rate, override rate, time-per-decision) and is treated as a conformity indicator under AI Act Art. 14.

AG-L4-06. Adversarial evaluation (red-teaming) is performed on a defined cadence and on every material change; results feed into risk management and incident-response readiness.

8.5 Layer 5 Controls, Audit and Evidence Governance

AG-L5-01. An audit envelope is generated for every agent session and retained per tier rules; envelopes are tamper-evident.

AG-L5-02. A decision envelope is generated for every decision within Art. 22 scope and is sufficient to support Art. 22 explanation, Art. 50 AI Act disclosure, and Art. 15 access responses.

AG-L5-03. ROPA records a compositional list of agent processing classes; every session instantiates one or more classes from the registered list; out-of-class activity is blocked or escalated.

AG-L5-04. DPIAs for agentic systems include the tool roster, memory architecture, delegation graph, and runtime enforcement architecture; they are updated on material changes.

AG-L5-05. Evidence of tier-by-tier retention compliance is produced on a defined cadence and made available on request.

AG-L5-06. Breach notification templates are pre-wired to the agent-specific indicators in § 7.7 to support 72-hour notification under Art. 33(1).

AG-L5-07. Third-party, independent audit of the runtime enforcement substrate is performed at least annually and on material change.

Catalog usage

The catalog is designed to be imported into a policy repository as-is. Organizations using the Governing Intelligence framework should map each control to the applicable layer in their existing AI Governance Stack instantiation, assign ownership, and reflect the controls in contractual artifacts (DPAs, supplier agreements, internal policies).

Part IX. Open Questions, Enforcement Gaps, and Research Agenda

This Part identifies issues where the analysis in this reference is incomplete, contested, or dependent on regulatory developments that have not yet occurred. Practitioners should expect to return to these issues as enforcement and standards activity matures.

9.1 The Controller Role in Multi-Agent Systems

Article 26 of the GDPR contemplates joint controllers who agree on their respective responsibilities. The article was drafted for stable bilateral relationships, not for multi-agent architectures where joint controllership may arise dynamically during a single session. The European Data Protection Board has not yet published guidance specific to agentic systems. Until it does, practitioners should document defensible defaults: designate the orchestrating agent's operator as the primary controller, maintain per-tool controller-role attestations in the tool registry, and be prepared to reclassify where the facts warrant.

9.2 The Scope of "Solely" Automated Decision-Making

Article 22 applies to decisions "based solely" on automated processing. Court of Justice case law (including SCHUFA, C-634/21) has narrowed the scope of human review that suffices to move a decision out of Art. 22. Agents are ripe for further case law because they can produce decision outputs with near-zero human friction. Organizations should track upcoming jurisprudence and err toward treating low-friction oversight as solely automated.

9.3 Memory Erasure and Model-Embedded Personal Data

Erasure from parametric memory (fine-tuned weights, pretrained models) remains a research frontier. Techniques such as machine unlearning, targeted gradient ascent, and differential privacy at training time are promising but not yet mature. Controllers should disclose the limits of erasure in privacy notices and should maintain credible retraining pathways as a backstop.

9.4 Evidence Standards for Runtime Enforcement

No consensus standard exists for evidence produced by runtime enforcement substrates. Auditors and regulators will need to evaluate policy-engine logs, audit envelopes, and decision envelopes without established assurance patterns. ISO/IEC 42006 (requirements for bodies auditing AI management systems) is a developing response; a complementary audit standard for runtime evidence is a plausible next step.

9.5 Cross-Framework Harmonization

The four frameworks in this reference overlap substantially but are not identical. Cross-framework audits today require redundant artifacts; integrated-assurance approaches are emerging but nascent. The author's research program includes work on a unified assurance schema covering GDPR DPIAs, AI Act technical documentation, NIST RMF profiles, and ISO/IEC 42001 evidence; that work will be reported in future publications.

9.6 Research Agenda Summary

The research program implied by this reference includes: (a) formal definition and measurement of agentic metrics suitable for regulator adoption; (b) machine-readable specifications for audit envelopes and decision envelopes; (c) machine-unlearning benchmarks for erasure compliance; (d) automated red-team corpora for agent-specific threats; (e) integrated assurance schemas across the four frameworks; and (f) economic analysis of runtime-enforcement investment returns for regulated enterprises.

Appendix A. Glossary

AI Agent. As defined in § 1.2, a machine-based system powered by one or more AI models that plans, uses tools, retains state, and produces effects beyond the return of a single generative output.

AI Governance Stack. The five-layer framework introduced in Governing Intelligence § 1.5: Data Governance (L1), Model Governance (L2), System Integration Governance (L3), Control and Monitoring Governance (L4), Audit and Evidence Governance (L5).

Audit Envelope. A structured record of an agent session sufficient to reconstruct the session's processing activities, tool calls, memory operations, and decisions.

Compound Regime. The combined effect of GDPR and AI Act obligations when agentic processing implicates both regulations simultaneously.

Decision Envelope. A structured record of a specific decision sufficient to support Art. 22 explanation, AI Act Art. 50 disclosure, and Art. 15 access responses.

Delegation. The routing of subtasks from a top-level agent to one or more sub-agents, each with potentially distinct prompts, tool permissions, and data access.

GPAI Model. General-purpose AI model as defined in AI Act Art. 3(63).

Memory Scoping. A runtime control that attaches scope identifiers to memory entries and evaluates scope compatibility at retrieval time.

Policy-as-Code. The expression of governance rules as executable predicates evaluated at runtime.

Runtime Enforcement. The implementation of governance obligations as controls that execute alongside the agent, rather than as policy documents or review gates applied pre-deployment.

Tool Gating. The restriction of tool invocation to an allowlisted subset based on tenant, user, and task context.

Appendix B. Selected References

Kenney, N. M. (2026a). Governing Intelligence: Law, Privacy, Security, and Compliance in the Age of Artificial Intelligence (1st ed.). Digital 520.

Kenney, N. M. (2026b). Runtime Enforcement of AI Governance. Digital 520 Working Paper Series.

Kenney, N. M. (2026c). Privacy Impact Assessment of Claude Opus 4.7. Digital 520 PIA Series.

Digital 520 (2026). Operationalizing Governing Intelligence. Webinar, recording available on request.

Regulation (EU) 2016/679 (General Data Protection Regulation), OJ L 119, 4.5.2016.

Regulation (EU) 2024/1689 (Artificial Intelligence Act), OJ L, 12.7.2024.

National Institute of Standards and Technology (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1.

National Institute of Standards and Technology (2024). Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile, NIST AI 600-1.

ISO/IEC 42001:2023, Information technology, Artificial intelligence, Management system.

European Data Protection Board (ongoing). Guidelines and recommendations relevant to AI processing.

Court of Justice of the European Union, C-634/21 SCHUFA, judgment of 7 December 2023.

OWASP (ongoing). Top 10 for Large Language Model Applications.