Governing the Agentic Enterprise: Applying EW-AiRM™ to Agentic AI and Multi-Agent Risk

What the Enterprise-Wide AI Risk Management (EW-AiRM™) framework is, how it governs autonomous and interacting AI agents across all three layers and the HAiPECR™ ethical overlay, and what firms must do now to manage the cumulative risk.

EW-AiRM™
by Human-Ai.Institute
Briefing v1.3· 1 September 2026
Framework baseline: v0.52

IN ONE PARAGRAPH

Agents act; models answer. The move from generative AI that produces outputs to agentic AI that executes multi-step tasks, calls tools, holds credentials and increasingly interacts with other agents changes the enterprise risk picture in kind, not merely in degree. Industry data through 2026 shows agent estates roughly doubling within quarters while governance, visibility and identity controls lag far behind. **EW-AiRM™** was designed for exactly this shift: its Strategic, Operational and Resilience layers, the **HAiPECR™** ethical overlay and the **Five Non-Negotiables** already treat autonomy, tool use, emergent multi-agent behaviour and provider-side change as first-class governance objects. This briefing sets out the framework, applies it to agentic and multi-agent risk, and closes with concrete moves and the free companion tools at **www.ewairm.com**.

EXECUTIVE SUMMARY

1 What is EW-AiRM™?

An open, additive framework that augments, never replaces, existing enterprise risk management. Three layers (Strategic: six pillars; Operational: the MIT AI Risk Repository as a living taxonomy engine; Resilience: eight AI Black Swan categories), the HAiPECR™ ethical overlay across all layers, and Five Non-Negotiables that no composite score can average away.
Its lineage runs from the Global Conduct Risk Paradigm (2015) and Universal Conduct Risk Paradigm (UNECE, 2017) through HAiPECR (2023, OECD-listed) and the UNECE Common Regulatory Arrangement on products with embedded AI (2024) to the Wiley Finance volume publishing 12 November 2026.

2 Agentic AI under EW-AiRM™

Autonomy does not change the architecture; it raises the stakes on every element of it. The deploy-at-all question (P1), named individual ownership (NN1, P5), tested human override within hours (NN3) and lifecycle monitoring at machine velocity (P6, KXIs) become the difference between a governed agent estate and an ungoverned one. HAiPECR™ gates every agent before go-live and at every material change, including changes the provider pushes on its own schedule.

3 Multi-agent risk under EW-AiRM™

Interacting agents produce collective behaviour that no single agent exhibits: EW-AiRM™ names this AI Black Swan 7, Multi-Agent Emergence, and the Operational Layer classifies it under MIT subdomain 7.6. Heterogeneous foundation models compound the problem: different providers, update cadences, safety postures and failure modes inside one workflow. The response is architectural: agent-level identity and least privilege, inter-agent observability, circuit breakers, and human override that works at the level of the system, not just the single agent.

4 What firms must do now?

Cumulative risk is managed by sequencing, not heroics: inventory every agent (including shadow agents), assign NN1 owners, gate through HAiPECR™, set tier-appropriate KXIs with trigger-action mechanisms, rehearse system-level override, and stress-test the eight Black Swan categories. The free browser-based companion tools (Interactive Cube, HAiPECR™ Pre-Deployment Filter, Assessment Tool Demo Lite, Risk<>Controls Master Map) operationalise each step today at **www.ewairm.com**.

THE BACKDROP IN NUMBERS · WHY THIS BRIEFING, NOW

88%
OF ORGANISATIONS REPORT AN AI-AGENT SECURITY INCIDENT IN THE PAST YEAR

~2×
GROWTH IN ENTERPRISE AGENT ESTATES WITHIN A SINGLE QUARTER

21%
HAVE RUNTIME VISIBILITY INTO WHAT THEIR AGENTS ARE DOING vs. 82 % EXECUTIVE CONFIDENCE IN EXISTING POLICIES

22%
TREAT AGENTS AS IDENTITY-BEARING ENTITIES;
46 % STILL RUN AGENTS ON SHARED API KEYS

Source: All four figures are paraphrased from Gravitee's State of AI Agent Security 2026 report (December 2025 wave, 900+ executives and practitioners; April 2026 update, 750 executives). The 88% is the December 2025 confirmed-or-suspected rate; the April 2026 wave reports 54% experienced-or-suspected and 34.9 % confirmed, a fall the authors attribute to under-reporting and detection gaps as fleets doubled. OWASP published its first Top 10 for Agentic Applications in December 2025. Directional survey findings, not audited measurement; they all point the same way: adoption has outrun governance.

"The first AI governance question is not how to deploy. It is whether to deploy." — EW-AiRM™ Governance Maxim

What is EW-AiRM™? Lineage, history and applicability

More than a decade of conduct-risk and AI-governance work, distilled into one open enterprise framework.

EW-AiRM™
by Human-Ai.Institute

EW-AiRM™, Enterprise-Wide AI Risk Management, is an open framework for governing artificial intelligence across an entire organisation. It is deliberately additive: it overlays AI-specific governance onto the ERM foundations a firm already runs (ISO 31000, COSO ERM, the Orange Book, the IIA Three Lines Model and their peers), so risk registers, committee charters and assurance cycles remain the backbone. Two strategic objectives anchor every decision the framework touches: **organisational survival** and **long-term sustainable benefit/profit**.

LINEAGE: BUILT ON A DECADE OF PUBLISHED GOVERNANCE INSTRUMENTS

The UCRP, presented at UNECE in February 2017 and published on the UNECE website, underpins HAiPECR's Conduct Risk dimension. The UNECE Common Regulatory Arrangement (CRA) on products/services with embedded AI, for which the framework's creator served as project lead, is the regulatory-side counterpart to EW-AiRM™ at the enterprise level. EW-AiRM™ is the culmination of this line of work, not a first draft.

THE ARCHITECTURE: THREE LAYERS, ONE ETHICAL OVERLAY, ONE IMMUTABLE FLOOR

LAYER 1 · STRATEGIC

Six governance pillars

The decisions made before and around deployment, from **P1 Strategic Alignment and Necessity Assessment** through to **P6 Through-the-Lifecycle Monitoring and Adaptability**. Each pillar carries a purpose, one core question and the failure it is designed to prevent, and turns evaluation into documented decisions: pause, go-ahead, or stage-gate.

LAYER 2 · OPERATIONAL

Grounded in evidence

Risk identification built on the **MIT AI Risk Repository** (MIT FutureTech): 7 risk domains and 24 subdomains drawn from **more than 1,700 categorised AI risks**, paired with the MIT AI Risk Mitigation Database of **831 evidence-based controls** in four quadrants, plus KXIs (Key X Indicators) as the monitoring currency.

LAYER 3 · RESILIENCE

Eight AI Black Swans

Low-probability, high-consequence events that risk registers miss, from emergent capabilities and foundation-model concentration through **multi-agent emergence** to the **quantum cryptographic transition**. Posture shifts from anticipation to absorption: **playbooks, stress testing, escalation, recovery.**

HAIPECR™ · THE ETHICAL OVERLAY ACROSS ALL THREE LAYERS

Originating in 2023 as the **Human-Ai Paradigm for Ethics, Conduct and Risk**, HAiPECR runs across every layer as a continuous ethical filter. Its seven dimensions, **Human oversight and autonomy · Accountability · inclusivity and participation · Protection of data, privacy and neural integrity · Ethics · Conduct risk · Rights and non-harm**, map to the ten core principles of the UNESCO Recommendation on the Ethics of AI (2021), and the instrument has been listed on the OECD AI Policy Observatory's tool catalogue since April 2023. A documented HAiPECR pre-deployment assessment is itself Non-Negotiable NN2.

APPLICABILITY: ANY ORGANISATION, PROPORTIONATE DEPTH

EW-AiRM™ scales by tier rather than by sector: **Core** (under 100 employees, 1 to 5 AI systems, mostly COTS or API deployments), **Standard** (100 to 1,000 employees, 5 to 25 systems) and **Full** (over 1,000 employees, more than 25 systems). Tiers change the depth, never the architecture, and organisations progress between tiers as exposure grows, not as ambition does. A **five-phase roadmap** (Assess, Enhance, Build, Implement, Sustain) carries a firm from first inventory to living governance. Whatever the tier, the **Five Non-Negotiables remain the immutable floor**: any one of them at a failing status forces an F-Critical posture, regardless of the composite score.

"Governance that looks right is not the same as governance that works." — EW-AiRM™ Governance Maxim

Why EW-AiRM™ is different from traditional ERM

Not a rival to enterprise risk management:
the AI-specific layer traditional ERM was never built to provide.

EW-AiRM™
by Human-Ai.Institute

Traditional ERM rests on assumptions that generative and agentic AI systematically break: that a validated system behaves the same way until someone changes it, that risks fit static categories, that monthly reporting cadences match the speed of change, and that root causes can be traced through human-readable logic. EW-AiRM™ does not discard ERM; it names precisely where ERM's assumptions fail against AI and supplies the augmentation for each failure.

DIMENSION	TRADITIONAL ERM ASSUMPTION	EW-AIRM™ AUGMENTATION
System behaviour	Deterministic: validated once, stable until changed.	Probabilistic and drifting: continuous behavioural monitoring; hallucination treated as a property of the architecture, not a failure of it.
Risk categories	Static operational, legal, reputational and market buckets.	A living taxonomy: the seven MIT domains and 24 subdomains, updated as evidence emerges, so one AI incident cutting across every legacy bucket still maps to a causal mechanism.
Cadence	Monthly reports, quarterly committees.	KXIs with cadence matched to operational velocity, near-real-time for high-risk services, each with a named owner and a defined trigger-action mechanism.
Assurance	Pre-deployment validation and periodic audit.	Layered, through-the-lifecycle assurance: output behaviour, calibration metrics, anomaly detection, and third-line testing of detection and response readiness.
Tail events	Bounded by historical precedent.	A dedicated Resilience Layer for eight named AI Black Swan categories: absorb what cannot be predicted.
Accountability	Committees and shared ownership.	NN1: a named individual owner for every deployed AI system, from deployment through decommissioning.
Ethics	Adjacent policy, periodically reviewed.	HAiPECR™ as a continuous filter across all layers, with a gated pre-deployment assessment (NN2) producing a documented record.

THE FIVE NON-NEGOTIABLES · THE ALWAYS-ON FILTER

NN1 Named individual owner for every deployed AI system. · **NN2** HAiPECR pre-deployment assessment for every new deployment. · **NN3** Tested, documented human override capability, operable typically within four hours or less. · **NN4** Explicit, documented acceptance of residual AI risk by an accountable person. · **NN5** AI incident reporting pathway with retaliation protection (UK PIDA 1998; EU Directive 2019/1937).

Any one of these at a failing status forces an F-Critical posture, regardless of the composite score.

ABOUT THE CREATOR

Prof. Markus Krebsz is the creator of EW-AiRM™ and HAiPECR™, and Founding Director of the Human-Ai.Institute and De-Risking Solutions Ltd. He served as AI Project Lead at UNECE Working Party 6, where he led the world's first Common Regulatory Arrangement on products with embedded AI (ECE/TRADE/486, 2024), and contributed to the EU AI Office expert community and the General-Purpose AI Code of Practice process.

He is a working-group member of IEEE P7007.1, a StandICT.eu Standards Maker, a ForHumanity Certified Auditor, Honorary Professor at the University of Stirling and Clinical Professor of Practice at Woxsen University, with more than two decades of board-level risk experience across banking and regulation.

THE HUMAN-AI.INSTITUTE

An independent Think/Do Tank

Convening multilateral and multi-stakeholder dialogue on AI governance, engaging with the UN, UNESCO, the OECD, IEEE, the EU Commission and governments worldwide.

Methodological home of EW- AiRM™ and a founding member of the UN University Global AI Network.

www.human-ai.institute

WHAT EW-AIRM™ IS NOT

Not a replacement for ERM, cybersecurity, legal, compliance or model-development practice: it aligns with and supports them. Not a promise to eliminate rare, high-impact surprises, and not a substitute for executive judgement. The full toolkit lives in the Wiley volume and at **www.ewairm.com**.

Applying EW-AiRM™ to agentic AI: all three layers

Agents act; models answer.
The same architecture governs both, but autonomy raises the stakes on every element of it.

EW-AiRM™
by Human-Ai.Institute

What changes with agency. An agentic system perceives, plans, calls tools, holds credentials and executes multi-step tasks with minimal per-step human review. Four properties distinguish it from the chat-style systems most governance was written for: **autonomy** (actions initiated without approval at each step), **privilege** (delegated system access that often exceeds any single human user's), **speed and scale** (thousands of actions before a human can respond) and **interconnection** (one compromised or confused component can corrupt a whole workflow chain).
None of this requires a new framework; it just requires the existing EW-AiRM™ architecture applied fully.

LAYER 1 · STRATEGIC: THE SIX PILLARS, READ FOR AGENTS

PILLAR	AGENTIC READING OF THE CORE QUESTION
P1 · **Strategic Alignment and Necessity**	Should this task be delegated to an autonomous agent at all? Automation enthusiasm is the most expensive strategic failure; the whether-to-deploy question precedes every how.
P2 · **Organisational Readiness**	Can the firm actually operate an agent estate: identity management, incident response and governance capacity sized to agents acting at machine speed, not to human-paced workflows?
P3 · **Data and Technological Maturity**	Will data pathways, tool interfaces and memory stores sustain reliable agent behaviour in production, with automated checks for drift and for poisoned or stale context?
P4 · **Risk Tolerance and Prioritisation**	Which actions may an agent take autonomously, which require human-in-the-loop approval, and which are prohibited outright? Thresholds belong in change control, playbooks and procurement terms.
P5 · **Governance and Accountability**	Who owns each deployed agent? A named individual with real override authority (NN1), not a committee, and not the vendor.
P6 · **Through-the-Lifecycle Monitoring**	How is behaviour monitored as the agent, its tools, its foundation model and its context all change, including provider-side updates pushed on the provider's schedule, not the firm's?

LAYER 2 · OPERATIONAL

The taxonomy already speaks agentic

Agentic failure modes map into the seven MIT domains without inventing new categories: excessive agency and scope creep under **AI System Safety, Failures, and Limitations** (including subdomain 7.6, Multi-agent risks), prompt injection and credential abuse under **Privacy and Security** and **Malicious Actors and Misuse**, automation over-reliance under **Human-Computer Interaction**. Each classified risk links to the 831-control mitigation database across quadrants Q1 to Q4, so controls attach to specific failure modes, not to systems as generic requirements. KXIs give agents a monitoring currency: tool-call anomaly rates, permission-scope drift, override frequency, model version currency.

LAYER 3 · RESILIENCE

Agents concentrate the Black Swans

Of the eight AI Black Swan categories, agents intensify at least five: **1 - Emergent capabilities** (behaviours absent in validation surfacing in production), **2 - Systemic foundation model concentration** (one provider-side change propagating through every agent built on it), **3 - Alignment failures** (an agent doing exactly what it was trained to do, at machine speed, where that diverges from intent), **5 - Adversarial breakthroughs** (injection techniques that outpace human-limited defences) and **7 - Multi- agent emergence**, treated in depth in Part 3. Resilience shifts posture from anticipation to absorption: rehearsed playbooks, circuit breakers and recovery arrangements assume an agent incident will occur.

HAIPECR™ APPLIED TO AGENTS · THE GATE BEFORE AUTONOMY

Every agent deployment passes the seven-dimension filter before go-live and again at material change reviews. For agents the dimensions bite hardest at **H** (is the human override real, tested and timely, or nominal?), **A** (a named individual accountable for actions the agent takes unsupervised), **P** (what the agent's memory accumulates over time, and who can see it), **C** (conduct risk is symmetric: the agent's conduct and the conduct of humans deploying, prompting and supervising it) and **R** (harms the agent could cause to parties who never consented to interact with it). The documented outcome, Proceed, Proceed with Conditions, or Halt Deployment, is exactly the record NN2 requires.

"Accountability cannot live in a committee: A named human being must own every deployed AI system."
— EW-AiRM™ Governance Maxim

Practical implications and proposed next steps

Two tracks: firms already running agents, and firms about to.
Both start with the same question.

EW-AiRM™
by Human-Ai.Institute

PRACTICAL IMPLICATIONS FOR THE ENTERPRISE

Identity is now a first-class risk object

Every agent needs a distinct, managed identity with least-privilege scoping and short-lived credentials. Agents granted broader access than their function requires is the most consistently reported failure pattern. Under EW-AiRM™ this is a P4/ P5 obligation with Q2 (Technical and Security) controls attached, and permission-scope drift becomes a standing KXI.

The perimeter moves to the tool call

For agents, the consequential moment is not the model output but the action: the API invoked, the record changed, the message sent. Runtime policy enforcement at the tool-call layer, with human-in-the-loop approval for designated high-impact actions, is where P4 risk-appetite thresholds become operational reality.

Shadow agents multiply silently

Employees and vendors stand up agents outside governance faster than registers update. The Assess phase inventory must actively hunt for unofficial agents, embedded agents inside SaaS products, and agent capabilities switched on by providers within tools the firm already licenses.

Provider cadence is your cadence

Foundation-model providers push updates on their schedule. A model version change can alter agent behaviour with no change on the firm's side: P6 monitoring, the model version currency KXI and post-update regression testing convert that external cadence into governed events rather than silent drift.

NEXT PROPOSED STEPS · MAPPED TO THE FIVE-PHASE ROADMAP

TRACK A · ALREADY DEPLOYING AGENTS

- **Assess:** inventory every agent, including shadow and vendor-embedded agents; map each against the six pillars; baseline-score the highest-risk agents first.
- **Enhance:** assign an NN1 named owner to every live agent this quarter; run a retrospective HAiPECR assessment on anything that went live without one.
- **Build:** define per-agent action tiers (autonomous / approval-gated / prohibited) and embed them in change control and procurement terms.
- **Implement:** stand up KXI dashboards for the highest-risk agents; test the NN3 human override end-to-end and time it.
- **Sustain:** rehearse an agent-incident playbook, including a provider-side model update scenario, at least annually.

TRACK B · PLANNING TO DEPLOY AGENTS

- **Assess:** answer P1 honestly: does this task need autonomy at all, and what does failure cost? Document the whether-to-deploy decision.
- **Enhance:** educate the board and risk committee on the seven MIT domains and the agentic reading of each pillar before the first agent ships.
- **Build:** run the HAiPECR Pre-Deployment Filter (NN2) on the candidate deployment; design identity, least privilege and override into the architecture, not after it.
- **Implement:** go live stage-gated: constrained scope, human-in-the-loop on consequential actions, expansion only on KXI evidence.
- **Sustain:** pre-agree the triggers that pause or roll back the agent, so escalation is automatic, not a debate.

THE AGENTIC HORIZON

Agentic systems that execute multi-step tasks with minimal human review will shrink today's human checkpoint. Governance must prepare now: tighter intervention controls, richer observability, stronger acceptance criteria.

Firms that cannot keep pace will struggle to compete; firms that chase capability without restraint will struggle to survive.
Many incidents labelled "user error" are better understood as collaborative misalignments between human cognitive bias and system design; reframing them keeps design teams accountable.

"Risk management, done well, does not slow innovation. It makes innovation sustainable."
— EW-AiRM™ Governance Maxim

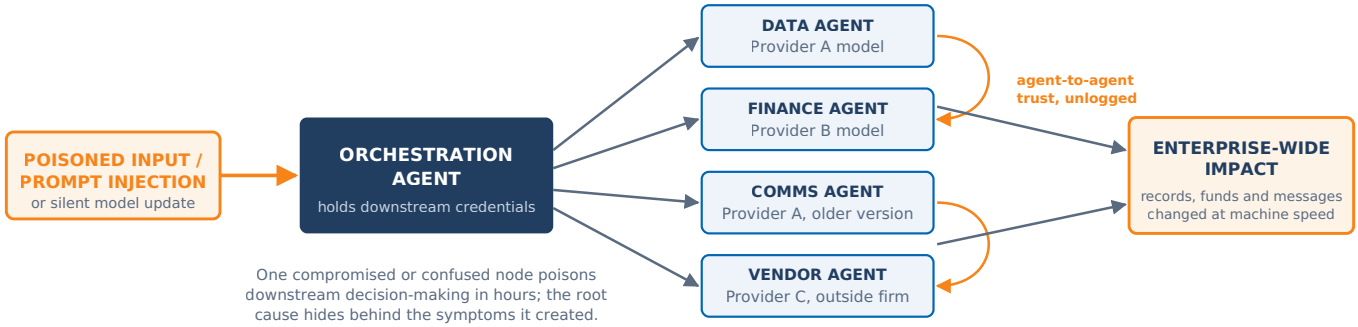
Multi-agent risk: when agents meet other agents

Collective behaviour that no single agent exhibits:
AI Black Swan 7, and MIT subdomain 7.6 - inside the enterprise.

EW-AiRM™
by Human-Ai.Institute

Single-agent governance is necessary but not sufficient. Once agents delegate to, negotiate with, or simply share infrastructure with other agents, whether the firm's own, a vendor's, or a counterparty's, the system acquires failure modes that exist only at the level of the interaction. EW-AiRM™ names this explicitly: **AI Black Swan 7, Multi-Agent Emergence**, in the Resilience Layer, classified operationally under **MIT subdomain 7.6, Multi-agent risks**. Autonomous agents interacting at scale produce collective behaviour, and failure modes, that no single agent exhibits: miscoordination, conflicting objectives, feedback loops, resource contention and cascading error propagation faster than human incident response can contain.

HOW A MULTI-AGENT CASCADE PROPAGATES



THE HETEROGENEOUS FOUNDATION-MODEL PROBLEM

Realistic enterprise agent estates are not built on one model. **Different agents run different foundation models, from different providers, at different versions, and increasingly interact with external agents the firm neither built nor governs.**
That heterogeneity creates four distinct challenges:

Divergent behaviour and safety postures

Each provider trains, aligns and guards its models differently. Two agents given the same instruction can interpret it differently; a hand-off that is safe within one provider's guardrails may be unsafe once it crosses to another's. No single-model assessment predicts the compound behaviour (AI Black Swan 4, multi-modal and compound effects, applies with full force).

Unsynchronised update cadences

Providers push updates independently. A workflow validated on Monday can involve three silently different models by Friday. Version currency must therefore be tracked per agent, and regression tested per interaction path, not per system.

Concentration hiding inside diversity

Many "different" agents resolve to the same handful of upstream providers (AI Black Swan 2). A provider-side outage or behavioural change becomes a correlated, estate-wide event precisely when the firm believed it had diversified.

Trust and attribution across boundaries

When agents authenticate to each other, credentials, context and instructions cross model and organisational boundaries. Without agent-level identity and complete inter-agent logging, the firm cannot attribute an action, reconstruct a cascade, or satisfy a supervisor asking who decided what.

AGENTS YOU DO NOT OWN

The hardest multi-agent exposure is often external: customer agents negotiating with the firm's agents, vendor agents embedded in licensed software, counterparty agents transacting through open protocols. The firm cannot govern their models, but it can govern the interface: authenticate every external agent, constrain what yours will disclose or accept, log every exchange, and treat agent-to-agent protocols as third-party dependencies under supply-chain controls (AI Black Swan 6) and procurement terms (P4, P5).

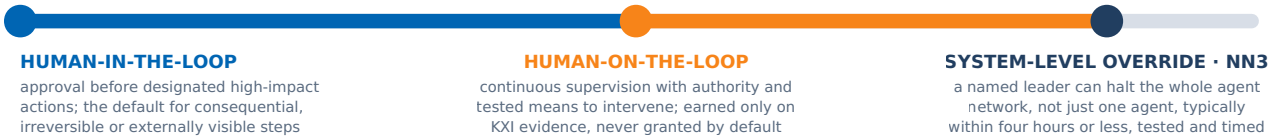
"The response to unknown unknowns is not anticipation, it is resilience." — EW-AiRM™ Governance Maxim

Human control, mitigants and the risk manager's playbook

Override at the level of the system, controls attached to failure modes, and how EW-AiRM™ equips the second line.

EW-AiRM™
by Human-Ai.Institute

HUMAN CONTROL: FROM PER-ACTION APPROVAL TO SYSTEM-LEVEL OVERRIDE



HUMAN-IN-THE-LOOP
approval before designated high-impact actions; the default for consequential, irreversible or externally visible steps

HUMAN-ON-THE-LOOP
continuous supervision with authority and tested means to intervene; earned only on KXI evidence, never granted by default

SYSTEM-LEVEL OVERRIDE · NN3
a named leader can halt the whole agent network, not just one agent, typically within four hours or less, tested and timed

EW-AiRM™ recognises two oversight modes, human-in-the-loop and human-on-the-loop, and requires the firm to choose deliberately per action class (P4), evidence the choice, and revisit it after incidents and model updates. The multi-agent twist is that override must work at the level of the **system**: pausing one agent while its peers keep transacting on its last outputs is not control.
Human override rates are themselves tracked as control indicators, not dismissed as residual variance, and a cascade that outruns the tested override window is, by definition, outside appetite.

MITIGANTS AND CONTROLS · MAPPED TO THE FOUR MIT QUADRANTS

QUADRANT	REPRESENTATIVE MULTI-AGENT CONTROLS
Q1 · Governance and Oversight	An agent-interaction map maintained as a governed artefact; NN1 ownership per agent plus a named owner for each cross-agent workflow; NN4 residual-risk acceptance at system level; vendor-agent terms in procurement.
Q2 · Technical and Security	Distinct managed identity per agent; least-privilege, short-lived, cryptographically bound credentials; no shared keys held by orchestrators; sandboxing; runtime policy enforcement at the tool-call layer; circuit breakers on inter-agent traffic.
Q3 · Operational Process	Rate limits and spend/action budgets per agent; staged autonomy expansion; regression testing of interaction paths after any provider-side update; rehearsed multi-agent incident playbooks including a full-stop drill.
Q4 · Transparency and Accountability	Complete, tamper-evident logging of inter-agent messages and tool calls; per-agent behavioural baselines; anomaly detection tuned to interaction effects and slow-building hazards; audit trails that let a human reconstruct who instructed whom.

HOW EW-AIRM™ HELPS THE RISK MANAGER CONFRONTED WITH ALL OF THIS

A shared language

The seven MIT domains, 24 subdomains and the eight Black Swan categories give the second line, the first line and the board one vocabulary for agentic exposure, so a deepfake, an injection and a cascade stop being forced into the same legacy bucket.

Evidence, not assertion

Every classified risk links to the 831-control mitigation database; every KXI has a named owner and a trigger-action mechanism. The risk manager can show a supervisor which control attaches to which failure mode, and what happens when a threshold trips.

A floor that cannot be argued away

The Five Non-Negotiables convert the hardest conversations into policy: no named owner, no deployment; no tested override, F-Critical posture. Composite scores cannot average away a missing control, however impressive the demo.

"A risk appetite statement without specific thresholds is a press release, not a governance instrument."
— EW-AiRM™ Governance Maxim

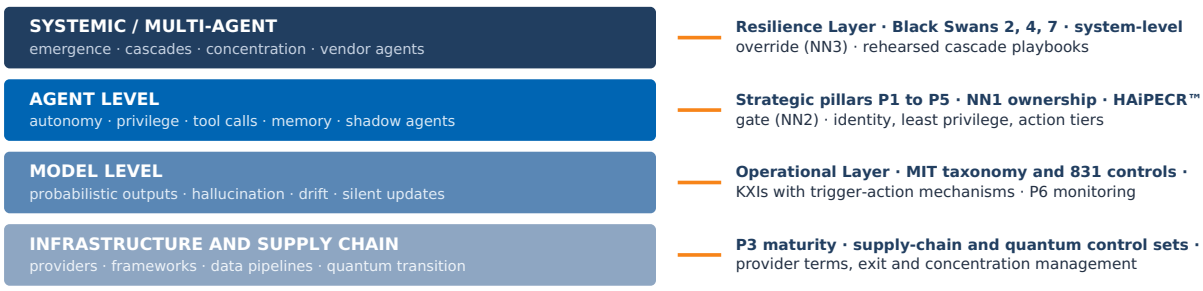
Bringing it all together: managing the cumulative risk

Single-agent, multi-agent and systemic exposures stack:
Governance must be sequenced, evidenced and rehearsed.

EW-AiRM™
by Human-Ai.Institute

What it all means. The agentic transition is not a future scenario; the incident data says it is a present operating condition. Risks now stack: model-level uncertainty (hallucination, drift) underneath agent-level exposure (autonomy, privilege, speed) underneath system-level emergence (interaction, contagion, concentration), all sitting on infrastructure and providers the firm does not control. Cumulative risk of this shape cannot be managed control by control or vendor by vendor. It needs one architecture that spans strategy, operations and resilience, keeps ethics continuously in the loop, and holds a floor that no business case can negotiate away. That is what EW- AiRM™ was built to be.

THE CUMULATIVE AI RISK STACK · AND WHERE EW-AIRM™ ANSWERS IT



WHAT FIRMS NEED TO DO · SEVEN MOVES, IN ORDER

- **1 · Inventory everything that acts.** Every agent, including shadow, vendor-embedded and externally interacting agents, into one governed register.
- **2 · Name the owners.** An NN1 individual per agent and per cross-agent workflow, with real override authority and accepted residual risk (NN4).
- **3 · Gate before autonomy.** HAiPECR™ pre-deployment assessment (NN2) for every new agent and at every material change, including provider-side updates.
- **4 · Set the action tiers.** Autonomous, approval-gated, prohibited: per agent, embedded in change control and procurement, evidenced under P4.
- **5 · Instrument at machine velocity.** KXIs per agent and per interaction path: tool-call anomalies, permission drift, override rates, model version currency.
- **6 · Rehearse the full stop.** Test the NN3 system-level override end-to-end, time it, and run a multi-agent cascade drill against the eight Black Swan categories.
- **7 · Report through what exists.** Feed AI-specific signals into the ERM reporting and escalation channels the firm already runs: augmentation, not a parallel bureaucracy.

HOW EW-AIRM™ HELPS

Each move above is a native EW-AiRM™ artefact, not an improvisation: the register and baseline scoring live in the Assess phase; ownership and gating are Non-Negotiables; action tiers are Pillar 4; instrumentation is the KXI engine; the full-stop drill is the Resilience Layer doing its job; and the reporting spine is the framework's founding design decision to augment, never replace, existing ERM. The next page shows the free tools that make each step executable this week.

THE FINAL WORD

The choice facing firms is not between agents and no agents; it is between governed autonomy and ungoverned autonomy. Governed autonomy compounds: every gated deployment, evidenced threshold and rehearsed override makes the next deployment safer and faster to approve. Ungoverned autonomy compounds too, in the other direction, one silent cascade at a time. The framework's promise is modest and specific: not that nothing will go wrong, but that when it does, the firm will know within hours, act within its tested override window, and explain itself to any supervisor, regulatory authority, counterparty or court with evidence rather than assertion.

"Hallucination is a property of the architecture, not a failure of it." — EW-AiRM™ Governance Maxim · and agents inherit the property, with tools attached

The tools on www.ewairm.com, deployed for agentic and multi-agent risk

Free, browser-based instruments. No sign-up, no install, and nothing leaves your browser (local install is also possible/available).

EW-AiRM™
by Human-Ai.Institute

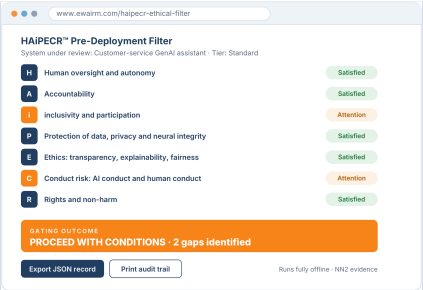
EW-AiRM™ Interactive Cube

The whole architecture, clickable, in a COSO ERM-style cube: three layers, six pillars, seven MIT domains, the HAiPECR™ overlay and all eight AI Black Swan categories. The fastest visual orientation for a board.

DEPLOY FOR AGENTIC / MULTI-AGENT RISK

Use it to brief the board on exactly where agentic exposure sits: click through Black Swan 7 (multi-agent emergence), Black Swan 2 (foundation-model concentration) and Pillar 6 before the next risk committee.

www.ewairm.com/ew-airm-cube



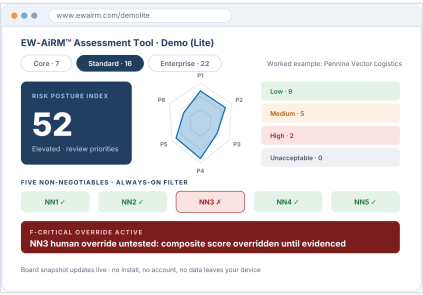
HAiPECR™ Pre-Deployment Ethical Filter

The seven-dimension overlay as a deployment gating tool: one focused decision, taken before go-live and at material change reviews, with outputs "Proceed", "Proceed with Conditions" or "Halt Deployment" and the specific gaps identified.

DEPLOY FOR AGENTIC / MULTI-AGENT RISK

Run it on every candidate agent and re-run it after any provider-side model update or scope expansion. The exported record is the NN2 evidence a supervisor will ask for, and dimension H forces the override question before autonomy is granted.

www.ewairm.com/haipecr-ethical-filter



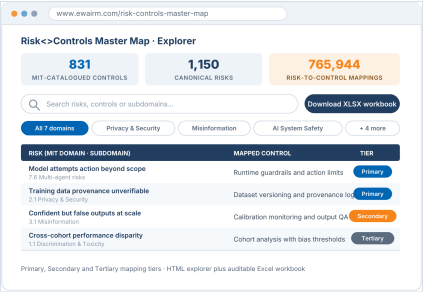
EW-AiRM™ Assessment Tool · Demo (Lite)

Pick a tier, answer the questions, and watch a board-level snapshot update live: Risk Posture Index, risk counts, Non-Negotiables status and the F-Critical override that no composite score can average away. Includes the Pennine Vector Logistics worked example for training teams.

DEPLOY FOR AGENTIC / MULTI-AGENT RISK

Baseline the firm's posture before and after the agent programme: an untested human override (NN3) shows up immediately as an F-Critical flag, which is precisely the conversation a scaling agent estate needs to force.

www.ewairm.com/demolite



Risk<>Controls Master Map

831 MIT-catalogued controls mapped to 1,150 canonical risks across the 7 domains and 24 subdomains, yielding 765,944 primary, secondary and tertiary mappings, with an interactive HTML explorer, an auditable Excel workbook and a methodology report.

DEPLOY FOR AGENTIC / MULTI-AGENT RISK

Filter to subdomain 7.6 (Multi-agent risks) and its neighbours to lift a ready-made, evidence-based control shortlist for agent identity, inter-agent logging and cascade containment straight into your register.

www.ewairm.com/risk-controls-master-map

TRY THE FRAMEWORK, NOT JUST READ ABOUT IT

The tools and the book are built from the same canonical reference base: when the framework updates, they update together. Tool previews above are illustrative renderings for print; run the live instruments at **www.ewairm.com/companiontools**.

Prof. Markus Krebsz & the wider EW-AiRM™ ecosystem

Where to go next: the book, the tools, the community and the conversation.

A NEW STANCE FOR ENTERPRISE RISK MANAGEMENT

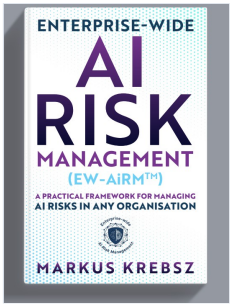
Relying solely on deterministic risk methods in the age of generative and agentic AI is itself a serious vulnerability: it stalls innovation while leaving the organisation exposed. EW-AiRM™ offers the alternative stance: augment rather than replace, ask whether before how, and build resilience over prediction. The framework exists to close the gap between governance that looks right and governance that works.

FURTHER EW-AIRM™ RESOURCES

Companion tools	The Interactive Cube, HAiPECR™ Filter, Demo (Lite) and Risk<>Controls Master Map (page 9), all free and browser-based: **www.ewairm.com/companiontools**
Website & video	Framework, tools, insights and the Knowledge-Bot: **www.ewairm.com** · YouTube: **youtube.com/@EW-AiRM**
HAiPECR on OECD.AI	Listed on the OECD AI Policy Observatory's catalogue of tools and metrics since April 2023: **oecd.ai/en/catalogue/tools/haipecr**
LinkedIn	**linkedin.com/company/the-human-ai-institute** and the thematic page **linkedin.com/company/enterprise-wide-ai-risk-management**

THE HUMAN-AI.INSTITUTE

An independent Think/Do Tank convening multilateral and multi-stakeholder dialogue on AI governance, engaging with the UN, UNESCO, the OECD, IEEE, the EU Commission and governments worldwide. Home of EW-AiRM™ and a founding member of the UN University Global AI Network. **www.human-ai.institute**



FORTHCOMING · WILEY FINANCE · 12 NOVEMBER 2026

Enterprise-Wide AI Risk Management (EW-AiRM™): A Practical Framework for Managing AI Risks in Any Organisation

The complete framework in one volume: the three layers, six pillars, HAiPECR™, the Five Non-Negotiables, the eight AI Black Swan categories and the implementation roadmap, with templates, KXI libraries, calibration matrices and worked examples throughout. ISBN 978-1-394-44647-6.

Pre-order: **www.amazon.co.uk/dp/1394446470**

CONTACT · CONSULTING & TRAINING

Email: contact@human-ai.institute
Phone: +44 (0) 79 85 065 045
Web: www.ewairm.com · www.human-ai.institute
Consulting & training: provided by De-Risking Solutions Ltd. and RiskAi.Ai

"Governance that looks right is not the same as governance that works." — EW-AiRM™ Governance Maxim

CITING THIS BRIEFING

Krebsz, M. (2026). *Governing the Agentic Enterprise: Applying EW-AiRM™ to Agentic AI and Multi-Agent Risk.* EW-AiRM™ Briefing, v1.3, 1 September 2026. Human-Ai.Institute / De-Risking Solutions Ltd. Available at www.ewairm.com. The framework baseline for this briefing is EW-AiRM™ manuscript v0.52; terminology follows the forthcoming Wiley Finance volume.