Use Case Risk Classification Framework

Classifying Agentic AI Deployments by Inherent Risk

Frédéric Verhelst · Version 1.1 · 16 June 2026

Purpose

The *Use Case Risk Classification Framework* scores a proposed agentic deployment by its inherent risk, independent of the organisation deploying it. It rates the use case on two axes, consequence (how bad the outcome if the agent errs) and likelihood (how likely an error reaches effect uncaught), and combines them into one profile: Low, Medium, High, or Safety-critical. The profile is one of two inputs to the deploy decision; the other is organisational readiness, from the *Agentic AI Readiness Assessment*. The board's risk appetite is applied on top. The framework is organisation-neutral: the same use case yields the same profile anywhere; what differs is the appetite the board sets and the readiness it holds.

Two axes, one matrix

Score the consequence axis by its highest single dimension across the six consequence dimensions; score the likelihood axis by its highest across the four likelihood dimensions; read the profile where they meet. The dimensions are on the worksheet overleaf.

Consequence ↓ / Likelihood →	Low	Med	High	Safety-critical
Safety-critical	**S-crit**	**S-crit**	**S-crit**	**S-crit**
High	**Med**	**High**	**High**	**S-crit**
Medium	**Low**	**Med**	**Med**	**High**
Low	**Low**	**Low**	**Med**	**Med**

A Safety-critical consequence is Safety-critical regardless of likelihood: a catastrophic outcome is treated as catastrophic however rare, the standard safety-engineering principle. Off the diagonal the cells carry real distinctions, a Medium consequence at Low likelihood is routine (Low), while the same consequence under constant errors is High. The output is a (consequence, likelihood) pair: the pair drives which controls to build; the overall profile drives the deployment gate.

Connecting to deployment

The profile is one leg of a three-way test: inherent risk here, organisational readiness from the *Agentic AI Readiness Assessment*, and the board's risk appetite, set as a fiduciary act. The preventive measures themselves, the constraints, contracts, and provenance that fire before an agent acts, are the *Agentic AI Capability Stack™*, owned and enforced under the *CDO Mandate*; this framework does not describe them, it sizes the risk so the CDO knows how much of that governance a deployment requires. The CDO, with the Chief Risk Officer and legal counsel, operationalises appetite into the *Minimum Maturity Deployment Thresholds*: the minimum readiness each profile requires on the capabilities it depends on. At the Deployment Gate the CDO checks the profile against readiness under those thresholds, and attests or halts. The relationship is monotonic, a higher profile demands higher readiness, with the enforcement that fires before an action kept at least as mature as the technology that enables it. This framework classifies; it does not set thresholds. Thresholds are organisation-, regulator-, and appetite-specific and carry legal exposure, so they are held internally, encoded as machine-readable guardrails (SHACL), and enforced at the gate; publishing them would create false comfort and legal risk.

The classification worksheet

Rate each dimension by the level that matches the proposed deployment. The consequence axis is the highest of its six dimensions; the likelihood axis the highest of its four. A use case Low on every dimension but one is set by that one. Combine the two on the matrix.

Consequence dimensions

Dimension	Low	Med	High	Safety-critical
Reversibility	Fully reversible at no cost.	Reversible at modest cost or delay.	Costly, slow, or only partly reversible.	Irreversible.
Blast radius	One task, one user.	One team or process.	A business unit or many customers.	Enterprise-wide, or the public and market.
Human impact and rights	No direct effect on individuals.	Affects experience or convenience; correctable.	Affects rights, access, or finances (credit, employment, services).	Affects physical safety, or fundamental rights at scale.
Data sensitivity	Public or non-sensitive data.	Internal or commercially sensitive data.	Personal data (PII), regulated financial, or trade secrets.	Special-category personal, or safety-critical operational data.
Regulatory exposure	No specific regulation.	General data-protection or consumer law.	Sectoral regulation, or binding EU AI Act obligations.	Conformity or licensing regime; breach can suspend operations.
Financial materiality	Below routine spend authority.	Within delegated budget.	Executive or CFO sign-off.	Board-level; could affect solvency.

Likelihood dimensions

Dimension	Low	Med	High	Safety-critical
Autonomy and oversight	Advisory; the system recommends, a human acts (in the loop).	Narrow pre-approved bounds; a human can intervene live (on the loop).	Autonomous within policy; reviewed after the fact (above the loop).	Autonomous, no effective oversight at decision time (out of the loop).
Velocity	Batch; hours or days to review.	Near-real-time; minutes.	Real-time; seconds.	Sub-second, high-frequency; no human-speed intervention.
Input trust and provenance	Internal, curated, validated at the Ingestion Gate (L2).	External but trusted sources, admitted at rest (L2) and validatable.	External or unvalidated data, or runtime data with limited checks.	External, untrusted, acted on at runtime (L8); prompt-injection exposure.
Scope and boundaries	Single data domain and discipline, within one team: one ontology, one owner.	Crosses data domains or disciplines within the organisation: several owners, one internal ontology to hold.	Spans many domains and disciplines enterprise-wide: federated ontology, multiple gates, many owners.	Crosses organisational boundaries (inter-company): no shared ontology, gate, or single accountable party.

Advisory, the Low autonomy level, is the pre-agentic baseline: the system recommends but a human acts. A purely advisory system is not an agent; the autonomy dimension measures how far a deployment has moved past that line.

Worked example

Take an agent that issues customer refunds autonomously against policy. On the consequence axis, reversibility is Medium (refunds reverse at a cost), blast radius is High (many customers), human impact is High (it affects customers' money and access), and data sensitivity is High (personal and financial), so the consequence axis is High. On the likelihood axis, autonomy is High (it acts within policy, reviewed after the fact), velocity is High (real-time), input trust is High (customer-submitted claims, external), and scope is Medium (it spans the customer, order, and financial domains), so the likelihood axis is High. High consequence at High likelihood reads High on the matrix, and the pair says build prevention and detection together. Had the same agent instead adjusted a physical process, the irreversible, physical consequence would set it Safety-critical regardless of how rare the error, the conservative rule in action.

About the author

Frédéric Verhelst, PhD (Applied Physics, TU Delft), works with boards and executive teams on the governance of trustworthy agentic AI. Twenty-five years across data, AI, and regulated industrial operations: external executive advisor on the Tyra Redevelopment FID (a Mærsk Oil joint venture with Shell and Chevron, later operated by TotalEnergies), then Head of Data Office at TotalEnergies EP Danmark, most recently bringing agentic AI into safety-critical maritime operations. Currently available for non-executive director (NED), board-advisory, and CDO appointments focused on corporate AI governance.

Canonical reference: the Agentic AI Capability Stack™ (v3). Companion artefacts: the Agentic AI Readiness Assessment, the CDO Mandate, the Board Briefing, the Vendor Coverage Diagnostic, and the CDO Role Specification. All artefacts at fredericverhelst.com/TOI-library. Licensed CC BY-ND 4.0 with attribution.

This framework provides an evaluation methodology. It is not regulatory advice and does not constitute an audit opinion. Risk appetite is set by the board; minimum maturity thresholds per risk profile are set by the organisation's Chief Data Officer with the Chief Risk Officer and legal counsel.