AI Maturity 2026 –

how key dimensions evolve as organizations advance across maturity levels

initiative for applied artificial intelligence

Dimension	L1: Experimenter	L2: Practitioner	L3: Professional	L4: Shaper
AI Ambition & Steering (strategy, governance, compliance, value)	AI ambition is emerging and largely framed as "let's start using AI/GenAI" to explore benefits and avoid falling behind. The enterprise mostly adopts what is readily available (standard tools, generic use cases) and learns what AI could mean for its context, without a sharp point of view on competitive advantage. Strategic decisions are driven by curiosity and early success stories; prioritization and risk stance are implicit rather than codified. Leadership acknowledges AI's importance but has not yet committed dedicated resources or integrated AI into strategic planning. Change management begins in a lightweight form through a first wave of communication and basic guidance to reduce uncertainty, inconsistent behavior, and shadow AI risk. *Failure mode:* AI ambition becomes a collection of experiments with no strategic throughline.	AI ambition becomes explicit and directional: the enterprise defines where it wants to play and sets priorities linked to business objectives (e.g., productivity, customer experience, risk reduction). Strategy increasingly reflects how the enterprise will leverage existing capabilities (including GenAI) in targeted domains-still primarily adopting the market's mainstream patterns, but with clearer choices and trade-offs. A baseline risk and compliance stance (incl. EU AI Act awareness) is translated into policies and decision forums, though application can vary across units. A senior executive has explicit accountability for AI. AI ambition is connected to business strategy, not treated as an IT initiative. Change management is professionalized for priority initiatives-moving from tool rollout to workflow integration and targeted enablement-yet adoption remains uneven across units. Responsible AI becomes more explicit via repeatable risk assessments and standardized documentation expectations, even if maturity varies by team. *Failure mode:* "Aspirational strategy" that does not translate into coherent strategic choices (too broad, too many priorities).	AI ambition is differentiated and measurable: the enterprise articulates how AI will create competitive advantage, not just efficiency gains, and ties this to concrete strategic themes and investment horizons. The company moves from adopting generic patterns to building distinctive capabilities (e.g., proprietary knowledge assets, domain-specific copilots/agents as strategic bets), guided by clear strategic governance and portfolio decision rights. Risk appetite and regulatory readiness are defined at strategy level-what the enterprise will and will not do, where human oversight is required, and how accountability is structured-enabling confident scaling without constant reinvention of policy. AI leadership has a formal seat in strategic decision-making with cross-functional authority. The executive team acts as role models. The enterprise begins to pursue AI-enabled business model innovation selectively (new offerings, data-enabled services, or differentiated propositions), beyond optimizing existing operations. Security and Responsible AI are treated as cross-cutting strategic enablers with defined expectations for oversight, documentation, and auditability. *Failure mode:* Strategy is strong but not adaptive-ambition lags behind model/market shifts and becomes outdated.	AI ambition is transformative and externally oriented: the enterprise aims to shape its industry by redefining customer experiences, operating models, or value chains around AI-first principles. Strategy is dynamic and learning-driven: the enterprise continuously repositions as model capabilities change, and it actively influences ecosystem direction (e.g., setting de facto standards, shaping regulatory discussions through informed participation, pioneering reference patterns). The risk posture is not merely compliant but strategic-trust, safety, and transparency are treated as competitive differentiators that enable faster adoption by customers, partners, and regulators. AI is inseparable from business strategy. Leadership continuously shapes the AI agenda and is recognized externally for AI thought leadership. AI and data underpin AI-enabled business model innovation at scale (new revenue streams and market positions), not only operational transformation. Responsible AI is a competitive advantage and a basis for external credibility and ecosystem leadership. *Failure mode:* Overreach-ambition to "shape the industry" outpaces the company's ability to sustain focus and credibility.
Use Cases (discovery, prioritization, product thinking, process integration)	Use cases emerge as a collection of ideas and pilots, often driven by local pain points or technology curiosity. GenAI pilots prioritize impressive demos and speed, while process integration and measurable outcomes remain secondary. The portfolio is essentially a backlog without a stable mechanism to decide what to scale or stop. *Failure mode:* "Demo success" is mistaken for business value, leading to pilot graveyards.	Use cases are prioritized using shared criteria (value, feasibility incl data readiness, risk), and ownership begins to resemble product ownership in selected areas. The enterprise learns to separate productivity GenAI from domain-critical applications that require grounding, controls, and change management. Reuse of patterns is limited, so scaling still feels like repeating the same work. *Failure mode:* Too many use cases start; too few reach sustained production impact.	Use cases are treated as products: roadmaps, releases, adoption metrics, and outcome KPIs are standard, and solutions are embedded into end-to-end processes. The enterprise chooses between ML, GenAI, and agentic approaches based on measurable fit and risk, not trend pressure. Successful patterns are reused across domains, and low-value use cases are retired deliberately. *Failure mode:* "Scale without focus"-many products exist, but value is diluted by weak prioritization discipline.	The enterprise continuously redesigns value streams around AI-first interaction and automation, using agentic orchestration where it is safe and value-positive. Use-case discovery becomes a strategic engine tied to long-term bets, and reuse is high through templates, reference patterns, and shared components. The organization sets a high internal bar for what "good AI products" look like and evolves it continuously. *Failure mode:* Over-automation-agents are pushed into workflows where constraints and human oversight are not yet sufficient.
Organization (operating model, roles, accountability)	AI work sits on top of existing roles, with unclear ownership and heavy dependence on a few individuals. Collaboration between business, IT, and risk/legal is episodic, creating slow decisions and rework when pilots touch sensitive areas. "Who runs it in production?" is often unanswered until the last minute. *Failure mode:* Key people become bottlenecks; progress stops when they move on.	A basic operating model emerges-often a small CoE plus business-led initiatives-with role definitions that make staffing possible beyond hero teams. Risk/legal/security engagement becomes more regular, and responsibilities for operating production AI start to be clarified. The model works in some areas but not yet consistently across the enterprise. *Failure mode:* CoE becomes a ticket queue-central team overload slows delivery and frustrates business units.	At this level, the enterprise has moved from a centralized setup to a hub-and-spoke operating model that scales delivery without losing coherence. A central AI Hub / CoE owns standards, reusable assets, governance enablement, and critical specialist capabilities (e.g., LLM/agent expertise, risk/legal/security interfaces), while strong Spokes in business domains own use cases and products end-to-end with clear accountability. In addition, the organization enables citizen development in a controlled way: business users can build low-risk automations or GenAI assistants within approved guardrails, with clear escalation paths and support from the hub. Cross-functional collaboration is dependable (business–IT–risk), and responsibilities for build vs run are explicit across domains. Change management capacity is explicitly staffed for priority initiatives, enabling workflow integration beyond tool rollout across multiple spokes. *Failure mode:* Hub-and-spoke exists on paper, but spokes are weak-everything funnels back to the hub, recreating a bottleneck.	At this level, the enterprise organizes AI as a core operating capability, not a support function. AI ownership is deeply embedded in every major unit: domains run their AI products and AI-first processes with clear end-to-end accountability (business outcomes, risk, and operational responsibility), while a powerful central function provides leverage where scale matters. That central function typically combines (1) an AI platform & enablement hub (standard stacks, model access, guardrails, reusable components), (2) a strategic transformation engine that can execute top-down, cross-domain AI-first redesign of critical value streams, and (3) a portfolio investment mechanism with material, often centralized budgets for a small number of enterprise-level AI bets. The operating model is designed to support multiple modes simultaneously: decentralized product teams for domain differentiation, centralized "tiger teams" for end-to-end transformation, and governed citizen-development channels for low-risk local automation-without fragmenting standards or accountability. Middle management routinely leads mixed human-and-agent teams across boundaries with clear delegation, oversight, and escalation practices. The organization continuously renews its change approach to prevent fatigue and maintain trust as autonomy increases, and it actively manages the long-term risk of core professional skill erosion through role design and governance routines. *Failure mode:* Organizational complexity overwhelms execution-too many parallel AI structures lead to unclear ownership and duplicated efforts.
Expertise (skills, enablement, knowledge management)	Skills are uneven and concentrated; learning happens through ad-hoc trainings, self-study, and informal peer support. Practical GenAI usage grows quickly, but shared standards for quality, safety, and evaluation are limited. Knowledge reuse is weak, so teams relearn the same lessons repeatedly. *Failure mode:* "Prompt folklore"-local tricks replace robust methods, producing inconsistent results.	The enterprise has moved from ad-hoc learning to role-based enablement for key groups (business, engineering, risk/legal), supported by playbooks and emerging communities of practice. Teams build baseline competence in choosing between traditional ML and GenAI approaches, and they understand fundamental safe-use practices (e.g., data handling, when human review is needed), even if depth varies by unit. Importantly at this level, C-level competence becomes "practitioner-grade": executives can distinguish hype from feasible impact, sponsor a focused portfolio, ask the right questions about data, risk, and value, and make informed trade-offs (speed vs. control, build vs. buy, central vs. decentralized). Executive understanding is sufficient to steer priorities and governance expectations, but not yet deep enough to consistently anticipate second-order effects of scaling GenAI/agents. C-level leaders begin to sponsor targeted change management for priority initiatives (workflow integration, communications, enablement), even if this is not yet systematic enterprise-wide. *Failure mode:* Executive sponsorship is symbolic-leaders endorse AI broadly but cannot make crisp investment and risk trade-offs.	Expertise becomes systematic and scalable: the enterprise runs a structured, frequently updated enablement system, tracks skill gaps, and codifies "how we build AI here" into reusable methods and patterns. Teams have practical mastery of MLOps/LLMOps fundamentals-evaluation, monitoring, versioning, and responsible-use mechanics-so delivery quality is no longer dependent on a few experts. At this level, C-level competence is decision-grade: executives can actively steer value creation, understand major GenAI/agent risk vectors, challenge providers and teams on evidence (not demos), and make durable investment choices across horizons. Crucially, mid-management competence becomes broad and operational: leaders of functions and value streams can translate strategy into roadmaps, redesign workflows, set adoption targets, manage incentives, and coach teams through the shift to AI-enabled work-turning AI from "projects" into a repeatable way of operating. The enterprise explicitly monitors and mitigates the risk of core professional skill erosion alongside AI augmentation, and middle managers develop leadership skills to run mixed human-and-agent teams with clear oversight and accountability. *Failure mode:* Strong top-level vision but weak middle layer-AI remains centralized or sporadic because managers cannot translate it into day-to-day execution and adoption.	he enterprise turns expertise into a compounding advantage: it develops differentiated methods, benchmarks, and reusable patterns that outperform generic best practices, and it scales these capabilities without relying on a small elite. Specialist depth exists in advanced domains (e.g., continuous evaluation at scale, agent safety engineering, domain-grounded GenAI, cost-performance optimization), and knowledge assets are continuously updated as models and risks evolve. At this level, C-level competence is transformation- and market-grade: executives can set AI-first direction for value streams, anticipate second-order effects (regulatory, workforce, trust, ecosystem), steer capital allocation across portfolios and horizons, and credibly represent the enterprise in strategic partnerships and external dialogues. Mid-management competence is institutional and ubiquitous: leaders across units can redesign operating models around AI, run disciplined experimentation with guardrails, manage productivity and risk trade-offs, and develop talent pipelines-so AI capability is embedded in how the enterprise learns and executes, not in isolated expert teams. The organization actively measures and manages long-term core professional skill erosion as autonomy increases, and it systematically equips managers to lead mixed human-and-agent teams (delegation design, oversight, escalation, performance management). *Failure mode:* Expertise becomes too "elite" and research-driven-cutting-edge competence exists, but it is not industrialized into everyday practices across the organization.
Culture (adoption, responsibility, change readiness)	Curiosity and hype coexist with uncertainty: some teams embrace AI, others mistrust it or fear workforce impact. GenAI use spreads informally, creating inconsistent norms and occasional risk incidents that reduce trust. Success stories exist, but adoption is not deliberately managed beyond early communication. *Failure mode:* Polarization-enthusiasts drive unsafe shortcuts while skeptics disengage completely.	Leadership increasingly normalizes AI usage and communicates the rationale, while beginning to address risk and job-impact concerns more transparently. Responsible-use norms appear in everyday behavior, and adoption grows in selected areas, though unevenly across the enterprise. AI becomes less "special," but not yet truly embedded in how work is done. *Failure mode:* Compliance backlash-one incident triggers restrictive policies that stall adoption.	AI-enabled ways of working are broadly accepted, and teams take responsibility for quality and responsible use rather than treating it as someone else's job. Change management is deliberate: champions exist, adoption is measured, and feedback loops continuously improve solutions. Trust grows because systems are reliable and oversight is clear, including for constrained agentic workflows. *Failure mode:* Adoption plateau-solutions are delivered, but incentives and workflow integration don't drive sustained usage	AI-first is a lived cultural norm: the enterprise does not merely "use AI," it reimagines how work is done-from decision-making and collaboration to how processes are designed, executed, and improved. Employees and leaders default to asking what can be automated, augmented, or agent-orchestrated, and they actively redesign roles and workflows around human judgment plus machine execution. Continuous improvement becomes faster and more empirical: teams run disciplined experimentation with guardrails, learn from telemetry and user feedback, and treat model and process updates as a normal operational rhythm. Responsible AI is deeply internalized as part of quality and professionalism, enabling high trust while moving quickly; people understand where autonomy is appropriate and where human oversight is essential, especially in agentic workflows. The organization continuously renews its change approach to prevent fatigue and maintain trust as autonomy increases, and it actively counters the cultural side effects of skill erosion (e.g., over-reliance on AI) through norms, coaching, and role expectations. *Failure mode:* "AI-first pressure" turns into burnout or loss of trust-pace increases without sufficient support, clarity, and safeguards for teams.
Data (strategy, quality, access, privacy)	Data is assembled per project, with friction around access, unclear ownership, and inconsistent quality. Pilots can be built, but scaling stalls when data constraints, privacy concerns, or missing provenance appear. For GenAI, knowledge sources are often unmanaged, leading to outdated or ungrounded outputs. *Failure mode:* "Data debt" surfaces late-teams discover they cannot operationalize because data is not fit for purpose.	Data ownership and access rules begin to stabilize, and reusable pipelines or data products appear for repeated needs. For GenAI, basic knowledge-source governance emerges (approved repositories for RAG, basic freshness and confidentiality practices). Improvements are real but uneven, often driven by a few high-profile use cases. *Failure mode:* Local optimization-each domain fixes its own data problems, creating fragmentation and duplication.	Data and knowledge are managed as enterprise foundations for AI: quality is measured, lineage is understood, and access is governed with auditable controls, so teams can scale solutions without repeatedly rebuilding data foundations. For GenAI, the enterprise operates disciplined knowledge sourcing-curated repositories, provenance expectations, freshness management, and systematic protection of sensitive content-so outputs are grounded and defensible. Context engineering for agentic AI is established: teams can reliably assemble task-relevant context (policies, process state, customer/account data, tool schemas, permissions, and recent actions) into structured, governed inputs that agents can use safely and effectively. This reduces hallucinations and unsafe actions by ensuring agents operate on approved, current, and minimal-necessary information. Data practices increasingly support Responsible AI requirements for traceability and auditability (e.g., provenance and access evidence). *Failure mode:* Knowledge sprawl-too many sources and weak curation degrade grounding quality and increase risk.	Data and knowledge operate as high-leverage products that enable AI-first processes across domains, often near real time, with strong provenance and governance that keep speed and assurance in balance. For GenAI, the enterprise maintains an actively managed knowledge layer-curated sources, freshness SLAs, citation/provenance expectations, and systematic handling of sensitive information-so outputs remain reliable at scale. Context engineering for agentic AI is industrialized: context is assembled dynamically from approved data products (state, policies, entitlements, tool schemas, prior actions) with strict minimization and traceability, ensuring agents act on the right information at the right time. In addition, the enterprise runs continuous input/output monitoring for agents from a data perspective: it tracks context quality, missing/expired sources, data leakage signals, action traces, and downstream effects to detect degradation, unsafe behavior patterns, and feedback-loop risks early. Responsible AI expectations (traceability, transparency, evidence) are supported by default through data and knowledge product design. *Failure mode:* "Context drift" – as processes and data change, agent context pipelines fall out of sync, leading to subtle but systemic errors that scale quickly.
Technology (platform, architecture, tooling, security by design)	Technology is fragmented: teams rely on standalone tools, trial accounts, and local setups, with limited standardization. Security controls for GenAI are basic and inconsistently applied, and integration with enterprise identity/logging is partial. Architecture is discussed, but a shared reference stack is missing. *Failure mode:* "Tool sprawl"-incompatible solutions proliferate, increasing risk and slowing consolidation later.	An approved baseline stack emerges: reference architecture, controlled access, and initial logging become standard for new solutions. Reusable components appear (e.g., RAG building blocks, basic guardrails), but teams still build variations and integration depth differs by unit. For early agentic use, allowlisted tools and coarse constraints may exist, though enforcement is inconsistent. *Failure mode:* Partial standardization-teams "comply" but bypass central patterns to move faster.	The platform reliably supports scaling: reusable components, strong observability across cost/quality/security, and security-by-design practices are institutionalized. For GenAI and agents, technical constraints-permissions, sandboxing, and action audit logs-are robust and standardized. Architecture enables integration into core systems without bespoke engineering for every use case. *Failure mode:* Platform bottleneck-central platform team cannot keep pace with demand, slowing innovation.	Technology becomes a strategic accelerator: golden paths and policy-as-code make secure delivery fast, and the platform evolves continuously based on enterprise needs. Multi-model orchestration and standardized agent tool layers enable AI-first automation with strong control and auditability. The platform supports rapid innovation without compromising security posture. *Failure mode:* Complexity creep-too many platform features reduce usability and drive shadow tooling again.
AI Ecosystem (vendors, models, partners, IP, cost/FinOps)	External tools and providers are adopted opportunistically, often without full transparency on data usage terms, IP implications, or cost exposure. GenAI spend can grow unexpectedly, and dependency on a single provider may form unintentionally. Ecosystem decisions are tactical and driven by speed rather than strategy, and approved enterprise alternatives are not yet established widely enough to curb shadow usage. *Failure mode:* Executives make provider decisions based on marketing promises rather than informed comparison of risk, cost, and strategic fit-creating early lock-in and avoidable exposure.	Vendor selection and contracting become more structured, with increasing focus on security, privacy, and regulatory alignment. GenAI usage and costs become visible enough to track, and basic spend controls and reporting begin. Provider choices start aligning with architecture and risk posture, though exit strategies and multi-provider resilience are still limited. *Failure mode:* Cost surprise-usage scales, but unit economics are unclear and budgets are exceeded.	Ecosystem management is proactive: SLAs, exit options, and multi-provider strategies are in place for critical capabilities, and IP/data protections are explicit. FinOps for GenAI is operationalized with budgets, guardrails, and optimization levers, enabling scaling without cost shocks. Partnerships accelerate delivery while critical know-how and differentiating assets remain controlled internally. Provider governance increasingly includes Responsible AI and security expectations (e.g., transparency, data usage terms, assurance artifacts), enabling credible scaling in regulated contexts. *Failure mode:* Governance slows partnerships-procurement/legal processes become so heavy that the enterprise misses technology windows and ends up with suboptimal stopgap solutions.	The enterprise orchestrates its ecosystem for sustained advantage: build/buy/partner choices are deliberate, and provider competition is used to improve performance, cost, and risk outcomes. The organization may influence market practices through reference implementations, standards participation, or selective contributions-while retaining strategic control over data and domain IP. Cost-value optimization becomes continuous and embedded in decisions. *Failure mode:* Ecosystem overreach-too many external bets distract from execution and internal capability building.
Execution (delivery, lifecycle, LLMOps/MLOps, evaluation, operations)	Delivery is exploratory and project-like: prototypes appear quickly, but productionization is rare and relies on individual effort. Testing and evaluation are informal, and monitoring is minimal, so reliability and repeatability remain low. GenAI is judged by qualitative demos, while agentic experiments are avoided or treated as novelty. *Failure mode:* "Pilot churn"-teams repeatedly rebuild prototypes with no pathway to stable operation.	The enterprise can bring selected AI solutions into production and keep them running, but practices differ across teams and scale is constrained. Basic delivery steps exist from scoping to deployment, and initial operational routines appear (incident handling, baseline monitoring). For GenAI, release checks are often manual and inconsistent; for agents, autonomy is limited and oversight is cautious. *Failure mode:* Reliability gap-solutions run, but quality issues and incidents erode trust and slow scaling.	Execution is standardized and measurable: evaluation, release gates, versioning, regression testing, monitoring, and continuous improvement are consistently applied. GenAI is operated with LLMOps discipline -offline/online evaluation, telemetry, prompt/RAG governance-while agentic systems run with clear oversight and action-level observability. Delivery scales across business units without sacrificing reliability. *Failure mode:* Metric overload-many signals exist, but teams lack clarity on which metrics trigger action.	Execution becomes continuous and largely automated: evaluation and monitoring operate as always-on systems, and improvement cycles run frequently with high confidence. The enterprise safely operates agentic workflows at scale using robust constraints, rapid rollback, and continuous learning loops that improve outcomes and reduce cost. Delivery excellence enables sustained AI-first transformation rather than periodic programs. *Failure mode:* Automation without governance-speed increases, but oversight fails to keep pace for higher-risk workflows.