

Enterprise Agentic AI Agile Framework v4.1

A Comprehensive “People and Process-First” Playbook

Release Date: May 2025 **License:** CC BY 4.0
Author: Devashish Saxena (devashishsaxena@gmail.com)

Purpose:

End-to-end operating model for conceiving, designing, testing, and governing enterprise-grade agentic AI systems. Assumes the prioritization of business use cases based on potential impact has already been done as a separate exercise.

The framework built is extensive and includes many potential activities at each phase. The intention is to provide practitioners with a comprehensive playbook from which they can adapt their approach based on the context of their specific use case, and the business environment at the enterprise.

Audience:

CDO, CIO, CTO, CAIO, CDAIO, Product & Engineering Leaders, Transformation PMOs.

Executive Summary

The landscape of artificial intelligence is rapidly evolving, with agentic AI systems moving beyond simple tasks to orchestrate complex workflows and business processes and deliver significant business value. **Embracing agentic AI isn't just an option; it's rapidly becoming table stakes for organizational effectiveness, offering a path to create an enduring advantage in a competitive landscape.** These systems, capable of operating with a degree of independence and adapting based on feedback, are poised to **transform core business functions**, from accelerating efficiency in research and development to automating tasks in procurement and dramatically improving customer experiences.

Unlocking this potential, however, requires a deliberate and structured approach. While many organizations face the challenge of moving GenAI pilots to production, and analysts predict a significant percentage of projects will be abandoned, this framework, the Enterprise Agentic AI Framework v4.1, provides the **structured operating model needed to overcome these hurdles.** It guides organizations through the complete lifecycle of identifying, building, and deploying agentic solutions that integrate securely with existing systems, manage context effectively (e.g., leveraging techniques like RAG), and scale reliably to production. This framework is designed to help you achieve **real, ambitious ROI** by focusing on tangible business outcomes. It is built on a foundation

of a **pluggable and adaptable architecture**, preparing the organization for future advancements in this rapidly changing space.

Crucially, this framework is built on a foundation of **robust risk management essential for systems capable of taking action**. Deploying agentic systems demands necessary guardrails to prevent costly errors and ensure trust. Agentic systems can be inconsistent and unreliable, making continuous testing and evaluation crucial. The framework mandates proactive measures including preemptive risk evaluations, enforcement mechanisms like sandboxing, and continuous observability for real-time monitoring. By incorporating human oversight and control points – a **"semi-agentic" design**– the framework directly mitigates risks, especially in high-stakes environments, recognizing that full autonomy is not yet universally trusted. Continuous testing and evaluation are integral to ensuring accuracy and performance and are framed as crucial **"intellectual property"** for competitive navigation of the AI landscape. Guardrails are non-optional and should be coded in, running in parallel to prevent issues like prompt injection and manage output in high-risk scenarios. The framework incorporates specific checkpoints, such as a production go/no-go review based on live performance data, override counts, user trust signals, and cost data.

At its core, the Enterprise Agentic AI Framework v4.1 is a **"People and Process-First" operating model**. It emphasizes the vital importance of scoping problems effectively by defining "jobs to be done", fostering the necessary cross-functional collaboration between domain experts and technical teams, and ensuring education and handholding for business units to drive adoption and realize value. It deliberately avoids relying solely on tools, instead focusing on establishing the processes, roles, and evaluations needed to build trust and competence across the organization, preparing the team to move fast and deal with the inherent ambiguity of this technology.

In summary, the Enterprise Agentic AI Framework v4.1 provides the **essential blueprint** for organizations to **responsibly and effectively harness the power of agentic AI**, transforming pilot projects into production-ready systems that deliver **measurable business impact** while proactively managing the associated risks.

Ensuring Trust in Agentic AI Systems

Given that agentic AI systems operate with a degree of autonomy and will often interact with real-world systems, data, and potentially critical decisions – it is critical that in enterprise applications there is a strong focus on:

- 1) **Security**: how is the system protected against malicious attacks (e.g. adversarial attacks, data poisoning, prompt injection), unauthorized access and data breaches.
- 2) **Reliability and Robustness**: how will the system operate consistently, accurately, predictably, and handle unexpected inputs or failures gracefully.
- 3) **Bias and Fairness**: how does the system mitigate unintended biases in data or algorithms that could lead to unfair or discriminatory outcomes.

- 4) **Transparency and Explainability:** how does the end user of such a system understand how an agent arrived at its decisions esp. in certain regulatory applications such as healthcare or financial.
- 5) **Data Privacy and Protection:** how does the system handle sensitive data such as PII.
- 6) **Accountability:** who is accountable and responsible for the outcomes of an agentic system? Which human will be held responsible? In traditional systems in the enterprise, IT is often held accountable for performance, reliability, robustness of a system – how does this evolve for agentic systems that are built on a non-deterministic foundation.
- 7) **Ethical Considerations:** Adherence to ethical guidelines will play a critical role esp. in use cases in health care e.g.

As an example an enterprise trust posture could be reflected as following. However, these could vary based on enterprise and/or use case specific trust needs.

- Security grade: **ISO 27001 mapped**, zero hard-coded secrets.
- Privacy: **PII redacted at RAG retrieval**; row-level ACL.
- Kill-switch SLA: **< 30 s** tested quarterly.
- Model lifecycle: registry with upgrade checklist.

KPI Dictionary (enterprise-agnostic)

Metric	Definition	Why it matters
North-Star KPI	Single headline outcome (revenue, risk, experience)	Aligns agentic system design to business value/impact
SSAT / NPS	Stakeholder-Satisfaction score or NPS	Proxy for adoption, quality, or end-user “love”
Autonomy %	Interactions fully handled by agent	Shows ROI realization
Unit Service Cost	OPEX per completed interaction	Cost baseline and forecast
Escalation Rate	% routed to human oversight	Balance safety vs autonomy
Latency p95	95th percentile end-to-end time	Experience service level objective (SLO)
Policy Violations	Guard-rail breaches per 1k calls	Ethics & compliance health

The Enterprise Agentic AI Framework: Executive Overview

This framework outlines the critical phases and considerations for successfully and responsibly developing, deploying, and operating Agentic AI systems at scale, ensuring alignment with business objectives and robust risk management.

Foundational Consideration: Organizational Structure & Collaboration

Success at the enterprise level in implementing agentic AI systems requires breaking down silos and fostering collaboration across Product, Engineering, Legal, Compliance, Ethics, and Security teams.

Consider structuring teams to provide horizontal capabilities (e.g., shared evaluation infrastructure, centralized guardrail policies) to ensure consistency and efficiency across multiple agentic initiatives. This avoids redundant efforts and facilitates standardized governance.

Phase-Gate Calendar (example 16-week pilot)

Week 0-4	Phase 0 →
Week 5	Mission Definition Gate
Week 6-9	Phase 2 → Cost-to-Serve Gate
Week 10-12	Phase 3 → Ethics Gate
Week 13-14	Phase 4 → Prod Go / No-Go Gate
Week 15-16	Hyper-care roll-out in pilot mode
Week 17++	Repeat Phases 2-4 in ongoing sprints evolving agents/tools etc.

Phase 0: Human-Centric Discovery

Purpose: Deeply understand the current state: who are the end users, who are the internal actors, what is the current process, how does it perform today, what works well, where are the friction points? Keep in mind the overall objective is to reimagine the process, build agentic AI based automation and drive the **impact** the business seeks, by *removing the friction points* for the end users and the internal actors – the humans.

Activity	Description	Key Questions	Key Outputs
----------	-------------	---------------	-------------

End user journey mapping	For customer facing use cases (e.g. customer support) define current end user journey and understand the different personas and especially where the current friction points are.	Who are the end users? What is their journey today? What do they like about the current journey? Where are their friction points?	End user journey map by phases for different personas: what tasks are performed, where are current friction points, change-impact matrix
Current state Process Mapping	Build a common visual baseline of the current business process(es)	Current end-to-end flow? Bottlenecks, decision points, hand-offs?	Swim-lane map, pain-point heat-map
Business/Internal Stakeholder & Role Analysis	With the process map in place, capture who (internally) touches each step and their incentives/KPIs	Who does what? KPIs, incentives, friction?	RACI of employee actors, current pain points, change-impact matrix
Baseline Metrics Capture	With roles understood, pull in baseline hard numbers, understand trends	Current “impact” metrics: revenue, NPS or outcome satisfaction and unit service cost	Baseline KPI dashboard (proof of impact), Data quality
Waste-to-Zero Workshop	Run a fast kaizen workshop – cross functional session designed to identify and eliminate every non-value-added step	Which manual steps can be eliminated before automation?	Simplified future-state flow with “zero waste”, waste log
Knowledge Codification	In cleaned up process, identify the fastest, simplest error-free sequence of steps that achieves the desired business outcome	What is the “golden path” SOP for this workflow?	Canonical SOP deck for prompt/agent design, decision trees
Feature Opportunity Sizing	Size the steps for agentic lift (speed, experience quality, risk reduction) using chance-impact or impact-feasibility scoring	Where could autonomous agents lift speed, experience quality, or lower risk?	Impact-feasibility matrix, prioritized use-case/feature backlog
Target-State Co-Design	With waste removed, SOPs codified, and opportunities ranked, design the future-state process that	How must the process evolve for autonomy and observability?	Future-state blueprint, re-engineered process(es), and workflows

	agents and humans will co-inhabit.		
--	---------------------------------------	--	--

Outcome: Bundle all Phase 0 artifacts into a **Human First Charter with baseline impact KPIs/metrics** that feeds Phase 1.

Phase 1. Strategic Agent Blueprint

Purpose: Turn the “Human First Charter” from Phase 0 into a crystal-clear, metrics-anchored direction for the first set of agents to be built – defining their purpose, scope and how they will deliver value whilst adhering to the ethical, risk and legal guard rails with clear escalation paths in place to humans as needed.

Key Activity	Description	Primary Roles	Key Outputs
Draft Agentic Epics	Convert each high-priority workflow into a single <i>Agentic Epic</i> statement: <ul style="list-style-type: none">• Role (Sales-Assist Agent)• Goal (qualify and route inbound leads)• Tools/Data (CRM API, pricing DB)• Constraints (privacy tier, SLA)• North-Star KPI (lead-conversion rate)• Optimization metric (cycle time)	AI Product Owner + Process Owner	Set of Epics—one per candidate agent
Define Success and Guard-Rails	<ul style="list-style-type: none">• Quantify North-Star KPI (revenue, cost, risk or customer/stakeholder experience) baseline vs target• Select 2-4 supporting/operating KPIs (cost-per-unit, SSAT, error rate)• Establish guard rails - document policy, legal, ethical, safety, brand/tone and performance/cost constraints (e.g., no PII spill)• Specify escalation rules (route to human) if confidence thresholds not met	AI Product Owner, Ethics Partner, Risk Lead	KPI & Guard-Rail Matrix (one row per KPI, one row per guard-rail; includes target, owner, data source). Escalation & Confidence Threshold Table (links each trigger to the Responsibility Contract owner)

Responsibility Contracts	For each Epic assign: <ul style="list-style-type: none"> • Agent Owner (accountable exec) • Human On-Call (real-time override) • Failure Action (auto-pause, reroute) 	Product Owner + Ops Lead	Updated Risk Register w/ contracts
Solution Architecture and Tech Feasibility Check	Align on high-level architecture (single agent vs multi-agent, RAG vs no-RAG, required tool integrations). Quick spike to confirm technical viability and token cost ballpark.	Agent Architect, Prompt Engineer, AgentOps Lead	Feasibility memo; rough infra sizing
Resource and Budget Alignment	Map required FTEs, sprint count, and infra spend. Ensure the 10-20-70 resource mix is still sensible (ensuring ongoing change/adoption activities)	Program PMO, CFO rep, Product Owner	Updated Cost-to-Serve model
Ethics and Alignment Pre-Check	Ethics Partner reviews Epics and guard-rails for bias, fairness, compliance. Flags items that must go through Ethics Gate later.	Ethics Partner	Pre-check sign-off or action items

Outcome: A formally approved **Strategic Agent Blueprint** comprising of agentic epic 1-pagers, target KPIs, key guard-rails, responsibility contracts, solution architecture, technical feasibility, resource, and budget ballparks.

Phase 2. Agent Architecture and Integration

Purpose: Turn the approved “Strategic Agent Blueprint” from Phase 1 into a detailed, build-ready plan (prompts, memory design, data/tool wiring, security guard rails, and a validated cost-to-serve forecast).

Key Activity	Description	Primary Roles	Key Outputs
Platform and Buy-vs-Build Decision	Evaluate commercial / OSS agent frameworks (e.g., CrewAI, LangGraph, AutoGen) vs bespoke option. Select the stack that meets guard-rails, latency, extensibility, and TCO targets.	Agent Architect, AgentOps Lead, Security	Platform decision note Risk acceptance if bespoke

High-level Architecture and Memory Design	Choose cognition pattern (single agent, planner-executor, multi-agent). Define memory tiers (short-term token window, episodic DB, long-term vector DB, audit log) and planning loop/flow.	Agent Architect, Data Engineer	Architecture diagram (planner, executor, memory tiers, tools, observability, security), Memory schema Planning loop spec (plan, act/execute, evaluate, record)
Tool and Data Integration Spec	List every external API, data product, or RAG corpus the agent will invoke. Document endpoints, auth, expected latency, cost limits, and observability hooks.	Prompt / Tooling Engineer, System SMEs	Toolchain map, Security data-flow diagram
Prompt and Policy Engineering	Draft prompt taxonomy - system prompt, role/persona prompt, task prompt, function/tool wrappers, fallback prompts, tone guide, policy prompts (PII, ethics constraints). Include inline tags for confidence thresholds and escalation cues.	Prompt Engineer, Ethics Partner	Prompt library (version controlled)
Reusable Asset Library Contribution	Store new prompts, wrappers, eval configs in a shared Cross-Pod repository; tag with metadata for searchability.	Cross-Pod Guild delegate	Updated enterprise asset catalog
Security and Compliance Design	Threat-model the agent: auth scopes, rate limits, data classification, audit fields. Map to guard-rails and SOC2 / ISO / HIPAA controls as needed.	Security Architect, Ethics Partner	Threat model matrix, Security requirements doc, Compliance mapping matrix, Ongoing security test plan

Evaluation Harness Set-up (a repeatable test case pipeline)	Build an automated test bed that objectively scores every new agent build against the KPIs and guard-rails defined in Phase 1 - so failures are caught prior to production. Configure open harnesses (agentbench, AutoGen-eval, custom test suites) aligned to KPIs & guard-rails. Draft baseline scenarios.	Simulation/Test Engineer, AgentOps Lead	Eval-config YAML / notebook (defines <i>what</i> to evaluate and <i>how</i> , then acts as the <i>executor</i> and <i>analyzer</i> presenting <i>insights</i> for review)
Prototype Spike and Cost Profiling	Build a thin vertical slice (happy path only) and run through evaluation harness to sample token, latency, and infra cost. Iteratively tune prompts / RAG chunking.	Architect, Prompt Eng, Ops	Cost-per-call range, Latency histogram
Cost-to-Serve Forecast and Stage-Gate Deck	Aggregate infra pricing, Ops FTE, 10-20-70 change mix. Verify data-quality readiness and produce “go / fix / defer” recommendation.	Product Owner, CFO rep, Ops Lead	Cost Forecast model, Stage-gate deck

Outcome: A formally approved **Agent Architecture and Integration** deck comprising of architecture, integration, cost forecast model, and data quality readiness.

Phase 3: Agent Behavioral Stress Testing

Purpose: Validate agent behavior against functional KPIs and guard-rails in a fully sandboxed, risk-tiered environment before any end-user exposure.

Key Activity	Description	Primary Roles	Key Outputs
Simulation Environment Boot-up	Spin up sandbox infra, load snapshot RAG, install mocks; seed synthetic user IDs.	Test Eng, DevOps	Sandbox environment
Risk-Tiered Test Plan	Map each tool/data call to Tier H/M/L; assign entry/exit gates	Test Eng, Security	Tiered test matrix

Synthetic and Edge-Case Dataset Build	Generate happy-path, edge, and stress datasets; include policy-violation probes.	Domain SME, Test Eng	tests/*.jsonl
Harness Execution and Metrics Capture	Run evaluation harness across all tiers; collect accuracy, policy, latency, cost.	AgentOps Lead	Raw run logs, metric CSV
Red-Team / Adversarial Blitz	Human red-teamers attempt jailbreak, PII extraction, cost abuse.	Red-Teamers, Ethics Partner	Red-team report, CVE list
Fallback-Path and Escalation Rehearsal	Force tool failures, low-confidence outputs; ensure escalation triggers fire.	Architect, Test Eng	Escalation drill report
Reinforcement Learning from Human Feedback (RLHF)	SMEs label 200–500 interaction pairs; tune model or prompt.	Prompt Eng, SME	Fine-tuned checkpoint / updated prompts
Safety Scorecard & Remediation Backlog	Consolidate results; tag blockers vs must-fix-later items.	Product Owner, Ethics Partner	Scorecard PDF; JIRA backlog
Ethics Gate Review	Present scorecard: sign-off, conditional go, or reject.	Ethics Board, Security, Product Owner	Formal Ethics approval

Outcome: Signed **Ethics-Gate approval plus a Safety Scorecard** showing accuracy, policy compliance, latency, and cost all within thresholds—clearing the way for limited human-feedback rollout.

Phase 4: Human Feedback and Refinement

Purpose: Expose the agent to real users in shadow or co-pilot mode, capture subjective trust signals, refine prompts/tools, and prove North-Star KPI lift without compromising safety.

Key Activity	Description	Primary Roles	Key Outputs
Shadow-Mode Launch	Agent runs in parallel to humans; outputs logged but not shown.	Ops Lead, Process Owner	Shadow log

Trust UX and Explainability Touchpoints	Inject confidence score, “why” button, tool call preview into UI.	Interaction Designer	Updated UI spec
User Education and Training Bursts	5-min explainer videos, FAQ, slack posts aligned to a holistic training/education plan	Change-Enablement, Process Owner	Training artefacts
Weekly Adoption Huddle	Process Owner, Ops, Product review SSAT, override count, North-Star trajectory.	Change-Enablement, Process Owner	Huddle minutes, tweak list
Prompt / Tool Refinement	Apply tweaks from logs + huddle; bump prompt version.	Prompt Eng, Architect	Updated prompts file
KPI Delta Assessment	Compare live shadow KPIs vs baseline; update Cost-to-Serve forecast if needed.	Product Analyst	Delta sheet
Prod Go / No-Go Review	Steering committee checks KPI deltas, user-trust signals, open risks; decide.	Exec Sponsor, Product, Security	Signed Go / rollback plan

Success factors: SSAT \geq baseline, override count trending down, trust cues understood, no unresolved Sev-1 issues.

Outcome: Production Go/No-Go decision backed by live SSAT, override, and cost data; updated prompt/tool version frozen for GA rollout.

Phase 5: Deployment, Operationalization and Continuous Alignment

Purpose: Gradually roll out full autonomy, operate the agent under defined SLOs, and maintain performance through continuous drift detection, value realization reviews, and model lifecycle governance.

Key Activity	Description	Primary Roles	Key Outputs
Gradual Roll-Out Plan	5 % → 25 % → 50 % → 100 % traffic over “n” weeks with rollback checkpoints.	AgentOps Lead, Process Owner	Roll-out program plan

Observability Dashboard Go-Live	Build observability dashboard using e.g. Grafana/Datadog monitoring: latency, cost, autonomy score, policy violations.	Ops, DevOps	Live dashboard URL
Alert & SLO Configuration	Define p95 latency, cost per interaction, violation count SLOs; hook to incident management systems (e.g. PagerDuty/Opsgenie).	Ops, Security	Runbook & alert rules
Drift Detection and Re-alignment Loop	Weekly run: eval harness on fresh data measuring agent accuracy, cost and tone on fresh production logs flagging statistically significant degradation (compare to baseline; auto-ticket if KPI drop > accepted threshold)	Ops, ML Eng	Drift report; retrain tickets
Kill-Switch and Escalation Drills	Quarterly test of manual and auto shutdown; post-mortem.	Ops, Ethics Partner	Drill report
Regular Ongoing (e.g. Quarterly) Value-Realization Review	Baseline vs live KPI gap; ROI update.	Product Owner, CFO rep, Steering Committee	NorthStar KPIs actual vs target trend
Underlying Base Model Lifecycle Management	Governance and tooling to version, monitor, upgrade, or deprecate the underlying LLM or fine-tuned checkpoints.	Simulation/Test Engineer, AgentOps Lead	Model registry entries (e.g. MLflow)

Outcome: Agent in **steady-state production with SLOs met**, quarterly ROI verified, and active processes in place for drift re-alignment and future model upgrades.

Stage-Gates

1. Strategic Agent Blueprint (after Phase 1)
2. Cost-to-Serve Forecast (after Design).
3. Ethics-Gate Approval (post Tier-2/3 Simulation).
4. Production Go / No-Go (post Feedback sprint).

RACI Heat-Map – Stage-Gates

Gate	Product	CISO	CFO	Process Owner	AgentOps Lead	Exec Sponsor
Strategic Agent Blueprint	A	C	I	R	I	A
Cost-to-Serve	A	C	A	R	C	I
Ethics Gate	C	A	I	R	C	A
Prod Go/No-Go	R	C	A	A	R	A

A=Approver, R=Responsible, C=Consult, I=Inform

Appendix

References & Lineage

1. PwC (2024) *Agentic AI: The New Frontier* (<https://www.pwc.com/m1/en/publications/documents/2024/agentic-ai-the-new-frontier-in-genai-an-executive-playbook.pdf>)
2. McKinsey (2025) *How COOs maximize operational impact from gen AI and agentic AI* (<https://www.mckinsey.com/capabilities/operations/our-insights/how-coos-maximize-operational-impact-from-gen-ai-and-agentic-ai>)
3. BCG (2025) *AI Agents as the All-Stars* (<https://www.bcg.com/publications/2025/how-ai-can-be-the-new-all-star-on-your-team>)
4. Agent Oriented Software Engineering (AOSE) literature (Wooldridge et al.)
5. OSS tool communities – LangChain, CrewAI, AutoGen, agentbench

Publication & Community Roadmap

Steps 0-9 as outlined in prior guidance, including license, repo structure, CHANGELOG, first community call to provide feedback, guidance and help evolve this base framework with the following guidelines:

- 1) Keep the framework focused on building agentic AI systems for the **enterprise**.
- 2) Carve out and evolve sector specific activities or even related frameworks e.g. healthcare sector may call for a very specific set of activities esp. around ethical, regulatory, and ethical compliance of any agentic AI system.

End of Playbook v4.1