Enterprise Agentic AI Agile Framework v5

A Comprehensive "*People and Process First*" Playbook

Executive Version for CxOs

Release Date: June 2025

Author: Devashish Saxena (devashishsaxena@gmail.com)

“That whole part of using Agentic AI to revolutionize the way we work inside companies, that’s just starting.”

Jensen Huang, CEO, **NVIDIA**, February 25, 2025

“Generative AI is going to reinvent virtually every customer experience we know, and enable altogether new ones about which we’ve only fantasized.”

Andy Jassy’s Letter to Shareholders, CEO, Amazon, April 10, 2025

Unlocking (*Agentic AI’s*) transformative power requires a strategic, focused approach based on a company’s overall business objectives. It also means that the 10–20–70 rule —10% of the effort should be focused on algorithms, 20% on technology and data, and the remaining 70% on people and processes—is more relevant than ever.

BCG, May 2025

Nearly eight in ten companies report using gen AI—yet just as many report no significant bottom-line impact. Think of it as the “gen AI paradox.”

McKinsey, April 2025

What is different about Agentic AI systems at the Enterprise level? Agentic AI systems are **non-deterministic**. This means that its outcomes vary in an unpredictable manner. This will require a "holistic re-haul" of how software is built and deployed in the enterprise. And most companies don’t understand the size of the change that they need to go through to capture meaningful impact from the deployment of agentic AI based systems.

Devashish Saxena’s LinkedIn Post, May 2025

Purpose

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

This framework primarily focuses on scaled deployment of agentic AI systems at the enterprise level. 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. As such this framework is designed to be modular and scalable, recognizing that agentic AI systems vary widely in complexity, risk, and context of use. While some projects may only need lightweight, LLM-assisted development, others require full-scale, cross-functional governance and rigorous testing.

To support teams in choosing the right path, consider the following Use Mode Spectrum:

Mode	When to Use	Characteristics
Lean XP Mode	Small, internal, low-risk agents	Simple user statements, LLM pair programming, fast iteration
Agile Pilot Mode	Focused pilots seeking measurable value	Select core phases (0, 2, 4), light team coordination
Enterprise Trust Mode	Regulated, scaled, or user-facing agents	Full framework: guardrails, ethics gate, observability, drift detection

Note: This document focuses entirely on the Enterprise Trust Mode.

It is intended for enterprise leaders and builders working in contexts where trust, compliance, impact measurement, and organizational integration are non-negotiable.

For experimentation or lean use cases, teams may selectively adapt portions of this framework — especially Phase 0 (Human-Centric Discovery), Agentic Epics, and KPI alignment — but are encouraged to evolve toward Enterprise Trust Mode for long-term sustainability.

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 Agile Framework v5, 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.

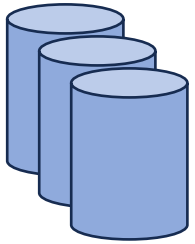
Crucially, this framework is built on a foundation of **robust risk management essential for systems capable of taking action**. Deploying agentic systems demands necessary guardrails to prevent costly errors and ensure trust. Agentic systems can be inconsistent and unreliable, making continuous testing 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 v5 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 v5 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.

Key Agentic AI Concepts



End User

Data
Access to internal or external, structured or unstructured data via e.g. RAG

AI Agent Core
The ”brain” or orchestrator of the system

Tools
Interact with other enterprise systems or external capabilities to perform actions via e.g. MCP

Memory

Allow agents to retain context over extended interactions and use past information for future decisions

LLM

The intelligence core that provides capabilities for reasoning, planning, and understanding/generating human-like text, which allows the agent to direct its own actions and perform complex tasks

Memory

Allow agents to retain context over extended interactions and use past information for future decisions

Human Oversight & Collaboration

”Semi-agentic" designs incorporate human-in-the-loop decision-making, allowing agents to seek human input when unsure or in high-stakes situations

Governance, Security & Reliability

Critical for responsible AI use comprising of non-optional guardrails, risk management and reliability

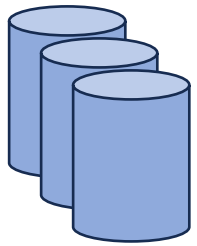
Observability: Evaluation & Monitoring

Observability is vital for understanding agent behavior in real-time, debugging problems, and iterating improvements via continuous evals

ERP

CRM

Pricing



Ensuring Trust in Agentic AI Systems

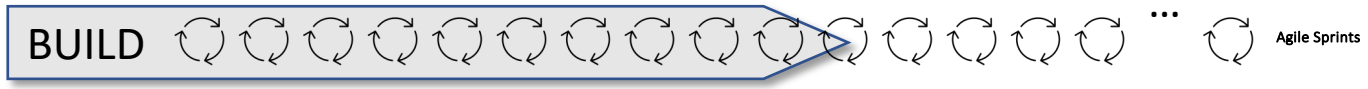
Given that agentic AI systems operate with a degree of autonomy and will often interact with real-world systems, data, and potentially critical decisions – it is critical that in enterprise applications there is a strong focus on:

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.

Enterprise Agentic AI Agile Framework v5



Phase 0: DISCOVERY Human-centric	Phase 1: DEFINE Agent Blueprint	Phase 2A: DESIGN Agent architecture, interactions & integrations	Phase 2B: DEVELOP Iteratively develop and validate	Phase 3: OPERATE Human feedback & iteration	Phase 4: DEPLOY Scaled rollout
Key Activities:					
• End-user journey mapping • Current state process map including internal stakeholder roles and responsibilities • Baseline KPIs including north star KPI • Target state process including updated stakeholder roles and responsibilities	• Identify & validate high impact business problems ← focus for *agentic* • Define target success KPIs (incl. north star) • Establish non-neg. risk guardrails • Build agentic epics • Preliminary – solution arch, tech feasibility, resource and budgets	• Agent prompt, policy engineering • Agent interaction design • Agent tool/data integration, memory design • Security & compliance design • Build automated evaluation harness(es) • Update Cost-to-serve forecast	• Develop agent components • Continuously test against automated eval harness(es) • Iterative refinement of agents • Human red-team adversarial blitz, escalation runs • Reinforcement learning / human feedback (RLHF) • Safety scorecard	• Shadow mode launch w/ humans in prod • User education and training bursts • Regular adoption huddle incl. KPI delta review • Agent/Tool refinement feedback	• Implement phased rollout strategy • Establish real-time monitoring and observability • Monitor for drift and value realization • Implement incident response and remediation • Collect, prioritize feedback & feed agentic roadmap for future iterations
Outcome:					
➢ Human-first, process focused charter w/ baseline KPIs (incl. north star)	➢ Agent blueprint (agentic epics, target KPIs, guardrails, solution arch, tech feasibility & est. budget)	➢ Agent interaction, architecture & integration definition	➢ Tested, robust & compliant agents ➢ Ethics approval & safety scorecard	➢ Production ready agents with Executive sign-off	➢ Agentic system in steady-state production w/ SLOs met, KPI validation, drift realignment & future LLM upgrades

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.

Key Activities:

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.

Current state process map including internal stakeholder roles and responsibilities:

Build a common visual baseline of the current business process(es) capturing clearly who (internally) touches each step and their incentives/KPIs.

Baseline KPIs including north star KPI:

Pull in baseline hard numbers, understand trends of current "impact" metrics – revenue, NPS or outcome satisfaction and unit service cost.

Target state process including updated stakeholder roles and responsibilities:

Design a reimagined future- state process that agents and humans will co-inhabit that removes wasteful steps and is built to take advantage of the strengths of agents and humans mapped to the biggest impact opportunities. This is a first step towards understanding the various agents that could be part of the target process.

Agent platform – buy vs build decision (optional):

Evaluate commercial / OSS agent frameworks (e.g. CrewAI, LangGraph, AutoGen) vs. build option. This may be seen as an enterprise decision or a by agentic product decision tied to specific guardrails, latency, extensibility and TCO goals.

Outcome:

Human-first, process focused charter w/ baseline KPIs (incl. north star)

Phase 1: **DEFINE** a strategic agent(ic) blueprint

Purpose:

This phase is about establishing the fundamental "why" and "what" for the agentic system, ensuring alignment on business value and identifying critical constraints from the outset. The “Human First Charter” from Phase 0 is turned 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 ethical, risk and legal guardrails with clear escalation paths in place for humans as needed.

Key Activities:

Identify & validate high impact business problems:

Clearly articulate the specific, meaningful business problem the agent will solve, focusing on areas where automation or agentic capabilities offer significant ROI or strategic advantage. Ensure this problem resonates across relevant stakeholders.

Define concrete, business-aligned target success KPIs (incl. north star):

Establish clear, measurable target metrics including the North-Star KPI (revenue, cost, risk or customer/stakeholder experience). Select 2-4 supporting/operating KPIs (e.g. cost-per-unit, SSAT, error rate). Determine how these metrics will be collected and measured.

Establish non-negotiable risk and ethics guardrails:

Define critical constraints established as guard rails clearly documenting policy, legal, ethical, safety, brand/tone and performance/cost constraints (e.g., no PII spill). This includes determining requirements for human oversight, intervention, and escalation rules, particularly for external or high-stakes applications where full-autonomy may not be trusted or used. Proactively identify security risks associated with agent actions and data handling. Align with ethics regarding bias, fairness and compliance.

Build agentic epics:

For each (potential) agent from target state process build an agentic epic 1-pager that includes: role (sales-assist agent), goal (qualify and route inbound leads), tools (CRM API), data (pricing DB), constraints (privacy tier, SLA), north star KPI (lead conversion rate), optimization metric (cycle time). For each epic assign agent owner (accountable exec), human-on-call (real time override), and failure action (auto-pause, reroute).

Build preliminary solution architecture, tech feasibility, required resource and budgets:

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. Map required FTEs, sprint counts and infra spend. Ensure 10-20-70 resource mix is still sensible (ensuring estimate effectively reflects ongoing change/adoption activities).

Outcome:

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

Phase 2a: **DESIGN** agent interactions, architecture, integrations

Purpose:

This phase designs the agentic system with built in quality, resilience and preliminary risk mitigation from the start. The Strategic Agent Blueprint from Phase 1 is turned into a detailed, build ready set of artifacts comprising of prompts, memory design, data and tool wiring, security guardrails, and a validated and approved budget forecast. Portions of the activities below could form the agile design sprints that run ahead of development agile sprints from Phase 2b.

Key Activities:

Agent 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. An enterprise agentic asset catalog would provide the ability to store and share prompts, wrappers, eval configs – tagged with metadata for searchability.

Agent interaction design:

> Design seamless and effective interactions between end users / humans and agents. This involves designing for both the agent's capabilities and the end user's needs, ensuring transparency, control, and a positive partnership. Exploit opportunities to go beyond text in visually engaging the end user.

Agent tool/data integration 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. List every external API, data product, or RAG corpus the agent will invoke. Document endpoints, auth, expected latency, cost limits, and observability hooks.

Security and compliance design:

> Define (ongoing) security and compliance plan by threat-modeling the agent: authorization scopes, rate limits, data classification, audit fields (for traceability and compliance). Map to guard-rails and SOC2 / ISO / HIPAA controls as needed.

Build automated evaluation harness(es):

> 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 (representative inputs/situations agents will be tested against comprising of test input – expected outcome pairs) into a test suite (YAML, JSONL, notebook).

Update cost-to-serve forecast:

> Build a thin vertical slice (happy path only) and run through evaluation harness to sample token usage (if using an LLM), latency (time/call), and infra cost (compute costs, API usage). Iteratively tune prompts / RAG chunking. Aggregate infra pricing, Ops FTE, 10-20-70 change mix. Verify data-quality readiness and produce “go / fix / defer” recommendation.

Outcome:

A formally approved Agent Interaction, Architecture and Integration deck comprising of interaction design, architecture, integration, cost forecast model, and data quality readiness.

Phase 2b: Iteratively **DEVELOP** and validate

Purpose:

This phase develops and rigorously validates the agentic system with formal governance review to ensure the agentic system is ready for broader deployment, confirming it meets all predefined criteria. This includes validating agent behavior against functional KPIs and guard-rails in a fully sandboxed, risk- tiered environment before any end-user exposure.

Key Activities:

Develop agent components:

Build all agentic components – agents, tools, data, memory, integrations, front end and more as defined.

Continuously test against automated eval harness(es):

Build a risk tiered test matrix assigning a risk tier (H/M/L) to every tool call, data source and action. Write test cases ensuring they cover common scenarios and edge case and adversarial scenarios. Auto run evaluation harness on a regular basis against all test cases producing key metrics (accuracy, latency, cost/interaction and policy compliance).

Human red-team adversarial blitz, escalation runs:

To test emergent LLM vulnerabilities that automated testing may miss assemble a human “red team” with the goal of jailbreaking or tricking agent into dangerous/unethical behavior. Validate that the safety net works with accurate escalation by agent and the triggering of the correct fallback action (log, pause, alert).

Reinforcement learning from human feedback (RLHF):

Use RLHF to improve agentic behavior by using human (SME) judgment ensuring that agent not only is correct but also helpful, matches enterprise style/tone, and adapted to the enterprise’s domain. RLHF recalibrates the agent within the context of the enterprise rather than general LLM basis.

Iterative refinement of agents:

Based on above iteratively refine the agents and/or other agentic components.

Safety scorecard:

Consolidate results and build an exec friendly Safety Scorecard that comprises functional accuracy, policy compliance, adversarial resistance, escalation handling, latency, and cost per call. Build a remediation backlog to address prioritized items.

Outcome:

Tested, robust & compliant agents that are ready for human-feedback deployment. Signed Ethics-Gate approval plus a Safety Scorecard showing accuracy, policy compliance, latency, and cost all within thresholds.

Phase 3: **OPERATE** with human feedback and iterate

Purpose:

Expose the agentic system to human users in shadow or co-pilot mode, capture subjective trust signals, refine prompts/tools, and prove North-Star KPI lift without compromising safety. This phase is critical in building real human user trust, and where agents evolve from prototype to production-readiness.

Key Activities:

Shadow mode launch with humans in production environment:

> Exposes the agents to real world data and flows without affecting production outcomes. The agent can run in the background watching and generating proposed responses allowing side-by-side comparisons with human decisions. Particular attention should be placed to trust UX and providing explainability hooks. Output logs identify trust cues, failure points and opportunities for tuning.

User education and training bursts:

> Prepare human users to understand, trust and correctly interact with the agentic system and avoid pilot rejection by creating and executing on training collateral. Focus on helping them understand the “why” tied to the target KPIs and the critical role they play in driving to success.

Regular adoption huddle incl. KPI delta review:

> Create a rapid learning loop between user experience and agent behavior tuning and ensure human-in-the-loop feedback continues through rollout.

Agent/Tool refinement feedback:

> Output from regular adoption huddles is a source for prioritized tasks/stories for future iterations.

Outcome:

Production ready agents with Executive sign-off. Production Go/No-Go decision backed by live SSAT, override, and cost data; updated prompt/tool version frozen for GA rollout.

Phase 4: **DEPLOY** with a scaled rollout

Purpose:

This phase focuses on the ongoing management of the agent in production, ensuring reliability, detecting and addressing issues, and continuously improving based on live performance and feedback. The goal of Phase 4 is to gradually roll out full autonomy, operate the agent under defined Service Level Objectives (SLOs), and maintain performance through continuous drift detection, value realization reviews, and model lifecycle governance. This phase handles the transition from testing/piloting (Phase 3) to production and ensures the system remains effective and reliable over time.

Key Activities:

Implement phased rollout strategy:

A crucial activity in this phase is executing a plan for increasing traffic to the agent in increments, such as 5% → 25% → 50% → 100% over a specified number of weeks. This process should include rollback checkpoints to allow for a safe retreat if issues arise.

Establish real-time monitoring and observability:

Build an observability dashboard to know “at real-time or near real-time how the agentic system is performing” (using e.g. Grafana/Datadog) monitoring latency, cost, autonomy score, policy violations. Execute on agreed upon remediation workflows and circuit breakers and ensure that all security, ethics and compliance guardrails are running effectively. Perform a regular test of manual and auto shutdown.

Monitor for drift and value realization:

Collecting and measuring the model's response through evaluations is ongoing. This includes understanding human preferences and building evaluations that capture these signals. Regular run of evaluation harness(es) on fresh production data measuring agent accuracy, cost and tone flagging any statistically significant degradation (compared to baseline or agreed KPI level. Conduct a regular and ongoing assessment of KPI vs baseline and targets.

Implement incident response and remediation:

Establish clear processes for detecting and responding to issues in production, including mechanisms like circuit breakers to limit negative impacts. Use monitoring data and incident analysis to inform the remediation backlog.

Collect, prioritize feedback & feed agentic roadmap for future iterations:

Gather feedback from live usage and users to inform ongoing development efforts. Use monitoring data, evaluation results, and feedback to drive continuous improvement cycles, iterating on the agent and its components (repeating activities from Build/Operate phases). Maintain continuous security monitoring and ensure ongoing compliance in the production environment.

Outcome:

Agentic AI system in steady-state production with SLOs met, quarterly ROI verified, and active processes in place for drift re-alignment and future model upgrades.

Key Roles Across Phases

Role	Short Description
Product Owner	Leads overall agent vision, prioritization, and value delivery; connects AI agents to business outcomes
Process Owner	Deep SME on the target workflow; ensures process redesign, adoption, and human-in-the-loop alignment
AI/Agent Architect	Designs agent architecture: planner, memory, tools, integrations, performance tuning
Prompt Engineer	Crafts and refines prompts, tool wrappers, and escalation logic to optimize agent behavior
AgentOps Lead	Owns day-to-day reliability, observability, guardrail compliance, and ongoing tuning after launch
Security Architect	Designs security posture: auth, rate limits, data protection, threat models
UX / UI Designer	Designs explainability and trust features in the agent-facing interface
Ethics Partner	Ensures fairness, transparency, and compliance with enterprise ethics and regulatory frameworks
Test Engineer	Builds evaluation harness, manages behavioral testing and safety validation
Data Engineer	Prepares data sources, RAG corpora, and supports memory tier management
Change Enablement Lead	Drives user training, change management, and adoption across impacted teams
Red Team Member	Conducts adversarial testing to uncover vulnerabilities and alignment risks
Program PMO	Oversees execution timelines, cross-phase gates, budget, and governance tracking
Executive Sponsor	Provides executive air cover, drives alignment with enterprise priorities, clears blockers
CFO / Finance Partner	Ensures financial viability of agent programs; reviews cost-to-serve and ROI metrics

Appendix

References:

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 Can Be the New All-Stars on Your Team (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
6. Deloitte (2021) AI governance for a responsible, safe AI-driven future (https://www2.deloitte.com/content/dam/Deloitte/us/Documents/risk/us-ai-governance- for-a-responsible-safe-ai-driven-future-final.pdf)
7. Moody's (2025) The rise of agentic AI in financial services: from automation to autonomy (https://www.moodys.com/web/en/us/creditview/blog/agentic-ai-in-financial-services.html)
8. Wang, Yue, & Chung, Sai Ho, Industrial Management & Data Systems (2022) Artificial intelligence in safety-critical systems: a systematic review (https://www.emerald.com/insight/content/doi/10.1108/imds-07-2021-0419/full/html)
9. Princeton University HAL: Holistic Agent Leaderboard (https://hal.cs.princeton.edu)
10. r/LocalLLaMa (2024) SomeOddCodeGuy My personal guide for developing software with AI assistance (https://www.reddit.com/r/LocalLLaMA/comments/1cvw3s5/my_personal_guide_for_develo ping_software_with_ai/)
11. Booz Allen (2024) Securing Artificial Intelligence (https://www.boozallen.com/content/dam/home/docs/ai/securing-ai.pdf)