

AI Comparison Center Handbook

Side-by-side decision guidance for insurance AI architecture and operations

CENTER	COMPARISONS
AI paradigms	Predictive AI vs Generative AI vs Agentic AI
Model adaptation	RAG vs fine-tuning
Operating disciplines	MLOps vs LLMOps vs AgentOps

AI Paradigms

Purpose	Predictive AI	Generative AI	Agentic AI
	Estimates a class, probability, score, ranking, or numeric outcome from observed data.	Creates or transforms text, images, audio, code, or other content from instructions and context.	Pursues a goal through planning, tool use, state, iteration, and action within defined authority.
Best fit	Predictive AI	Generative AI	Agentic AI
	Stable, bounded prediction problems with measurable ground truth.	Drafting, summarization, extraction, conversational assistance, and knowledge synthesis.	Multi-step workflows requiring coordination across systems and conditional action.
Insurance examples	Predictive AI	Generative AI	Agentic AI
	Severity estimation, fraud propensity, triage, churn, risk scoring, recovery likelihood.	Claim summaries, correspondence drafts, policy Q&A, document extraction, adjuster assistance.	Claim intake orchestration, evidence gathering, task routing, recovery workflows, service operations.
Output	Predictive AI	Generative AI	Agentic AI
	Score, class, probability, forecast, or rank.	Natural-language or multimodal content; sometimes structured output.	A completed workflow, tool actions, state changes, communications, and evidence.
Knowledge / change	Predictive AI	Generative AI	Agentic AI
	Learned statistical relationships in training data and engineered features.	Foundation-model training plus prompt context, retrieval, tools, or fine-tuning.	Models plus instructions, retrieval, memory, tools, policies, and current system state.
Evaluation	Predictive AI	Generative AI	Agentic AI
	Discrimination, calibration, error, stability, fairness, and business outcome.	Grounding, material factual accuracy, instruction following, usefulness, safety, and consistency.	Task success, action correctness, authority compliance, side effects, recovery, cost, and human rescue.
Operations	Predictive AI	Generative AI	Agentic AI
	Feature pipelines, training, registry, deployment, monitoring, retraining, rollback.	Prompt/version control, retrieval, evaluations, safety filters, cost/latency, monitoring.	Run traces, tool permissions, approvals, state, retries, limits, observability, rollback, incidents.
Key risks	Predictive AI	Generative AI	Agentic AI
	Bias, proxy discrimination, drift, data leakage, opacity, false positives/negatives.	Unsupported content, prompt injection, sensitive data disclosure, inconsistency, overreliance.	Unauthorized action, compounding errors, loops, hidden side effects, persistence, cascading harm.
Choose when	Predictive AI	Generative AI	Agentic AI
	Choose when the primary question is what is likely to happen or which case belongs in a defined class.	Choose when the primary objective is to create, transform, explain, or converse.	Choose only when autonomous or semi-autonomous action creates enough value to justify stronger controls.

Model Adaptation

Purpose	RAG	Fine-tuning
	Retrieves selected external information at run time and provides it to the model as context.	Updates model weights using curated examples to change learned behavior, style, format, or task performance.
Best fit	RAG	Fine-tuning
	Knowledge that changes, must be cited, is proprietary, or varies by jurisdiction or policy version.	Repeated stable behaviors that prompting cannot achieve reliably and for which quality training examples exist.
Insurance examples	RAG	Fine-tuning
	Policy and procedure Q&A, claim guidance, regulatory research, approved knowledge assistants.	Specialized classification, consistent structured extraction, tone or format, domain task adaptation.
Output	RAG	Fine-tuning
	Model output grounded in retrieved passages; the base model weights do not change.	A customized model that behaves differently without retrieving the training examples at run time.
Knowledge / change	RAG	Fine-tuning
	Indexes and retrieves approved sources at query time.	Training examples influence weights; not a dependable database for current facts.
Evaluation	RAG	Fine-tuning
	Retrieval recall/precision, source authority, freshness, grounding, citation fidelity, answer quality.	Holdout task performance, generalization, regression, safety, fairness, memorization, and drift.
Operations	RAG	Fine-tuning
	Ingestion, permissions, indexing, retrieval, source governance, monitoring, cache and version control.	Training-data governance, experiments, registry, evaluation gates, deployment, rollback, monitoring.
Key risks	RAG	Fine-tuning
	Stale or poisoned sources, wrong authority, access leakage, poor retrieval, indirect prompt injection.	Memorization, hidden data defects, bias, regression, stale facts, cost, vendor or base-model change.
Choose when	RAG	Fine-tuning
	Choose when the problem is getting current or proprietary facts into the model with traceable sources.	Choose when the needed change is stable behavior and sufficient governed examples justify training.

Operating Disciplines

Purpose	MLOps	LLMOps	AgentOps
	Operational practices for building, deploying, monitoring, and retraining machine-learning systems.	Extends operational discipline to LLM applications, prompts, retrieval, models, evaluations, safety, cost, and latency.	Extends LLM operations to goal-directed agents, tool calls, state, authority, handoffs, and long-running execution.
Best fit	MLOps	LLMOps	AgentOps
	Predictive models with data and feature pipelines, measurable targets, and repeatable training.	Generative AI applications whose behavior depends on model, prompt, context, retrieval, and validators.	Agents that take actions or coordinate multi-step work across systems.
Insurance examples	MLOps	LLMOps	AgentOps
	Pricing, underwriting, severity, fraud, triage, retention, and forecasting models.	Knowledge assistants, summaries, drafting, extraction, conversational service, document review.	Tool-enabled claim assistants, workflow agents, payment or communication orchestration, multi-agent processes.
Output	MLOps	LLMOps	AgentOps
	Reliable model versions and prediction services.	Versioned LLM application configurations and evaluated responses.	Governed runs, actions, traces, state transitions, approvals, and recoverable outcomes.
Knowledge / change	MLOps	LLMOps	AgentOps
	Training data, features, labels, code, model artifacts, and metadata.	Prompts, model versions, retrieved context, evaluation sets, policies, and feedback.	LLM context plus tools, credentials, memory, policies, workflow state, and downstream systems.
Evaluation	MLOps	LLMOps	AgentOps
	Offline/online model metrics, calibration, drift, data quality, business outcomes, and fairness.	Grounding, factuality, instruction following, usefulness, safety, robustness, cost, and latency.	End-to-end task success, tool/action correctness, policy compliance, loops, side effects, recovery, and rescue.
Operations	MLOps	LLMOps	AgentOps
	CI/CD/CT, feature store, registry, lineage, serving, monitoring, retraining, approvals.	Prompt registry, evaluation pipelines, retrieval observability, model routing, safety, tracing, monitoring.	Run tracing, sandboxing, identity, least privilege, approvals, budgets, retries, state, pause, rollback, incidents.
Key risks	MLOps	LLMOps	AgentOps
	Training-serving skew, drift, leakage, weak lineage, unstable retraining, ungoverned models.	Non-determinism, model/provider drift, prompt injection, unsupported output, data exposure, cost volatility.	Excess authority, cascading failure, persistence, hidden actions, unreliable recovery, cross-system impact.
Choose when	MLOps	LLMOps	AgentOps
	Use as the base operating discipline for predictive ML systems.	Use when production behavior depends materially on an LLM application stack.	Use in addition to LLMOps when the system plans, uses tools, changes state, or delegates work.

Decision Worksheet

Question	Decision / evidence
Insurance objective and affected decision	
Required output and knowledge freshness	
System action authority and reversibility	
Data, source, and training evidence	
Evaluation and counter-metrics	
Operating discipline and ownership	
Human oversight and escalation	
Risk tier, residual risk, and approval	

Sources

NIST AI RMF

NIST Generative AI Profile

Google Cloud MLOps architecture

OpenAI practical guide to building agents