Multi-Agent Design

Use role-based agents instead of one omnipotent agent.

Agent	Responsibility
Planner	Break down tasks
Researcher	Retrieval
Executor	Tool use
Verifier	Check outputs
Safety Agent	Policy checks
Finance Agent	Budget controls
Audit Agent	Logging

Orchestration Patterns

- Manager-worker — a coordinator agent dispatches to specialist workers and merges results.
- Task graph / stateful graph orchestration — the workflow is modeled as a directed (often cyclic) graph with conditional branching, persistent checkpoints, and interruptible human-in-the-loop points, rather than a fixed linear chain.
- Shared memory store — agents coordinate through a common memory layer instead of direct message passing.

Open problem: Byzantine fault tolerance in adversarial multi-agent settings remains an unresolved research area — don't assume agent-reported results are trustworthy without independent verification in high-stakes settings.

Model Routing / Cost Optimization

Use the cheapest model that can reliably complete the task.

Tasks	Model Tier
Classification	Small
Summaries	Small / Medium
Planning	Medium / Large
Coding	Specialist
Verification	Large
Legal / Compliance	Premium deterministic

Principle: Intelligence should scale with difficulty. Many current model families additionally expose configurable reasoning effort as a routing dimension in its own right — treat effort level as a tunable dimension alongside model size.

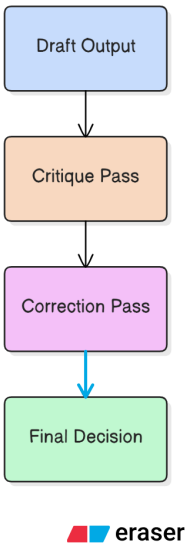
Reflection / Critique Loops

Use secondary reasoning for high-stakes decisions.

Trigger Conditions

- Low confidence
- High risk
- Contradictory evidence
- Large transaction
- Compliance-sensitive request

Reflection Flow



Human-in-the-Loop Design

Humans should be inserted intelligently.

Human Role	Tasks
Reviewer	High-risk decisions
Supervisor	Real-time override
Trainer	Correct outputs
Auditor	Compliance review

Confidence Handoff: Escalate when confidence < 0.85–0.90

Revision #1
Created 2026-08-24 12:07:38 UTC by Aisha M. Nur
Updated 2026-08-24 12:08:18 UTC by Aisha M. Nur