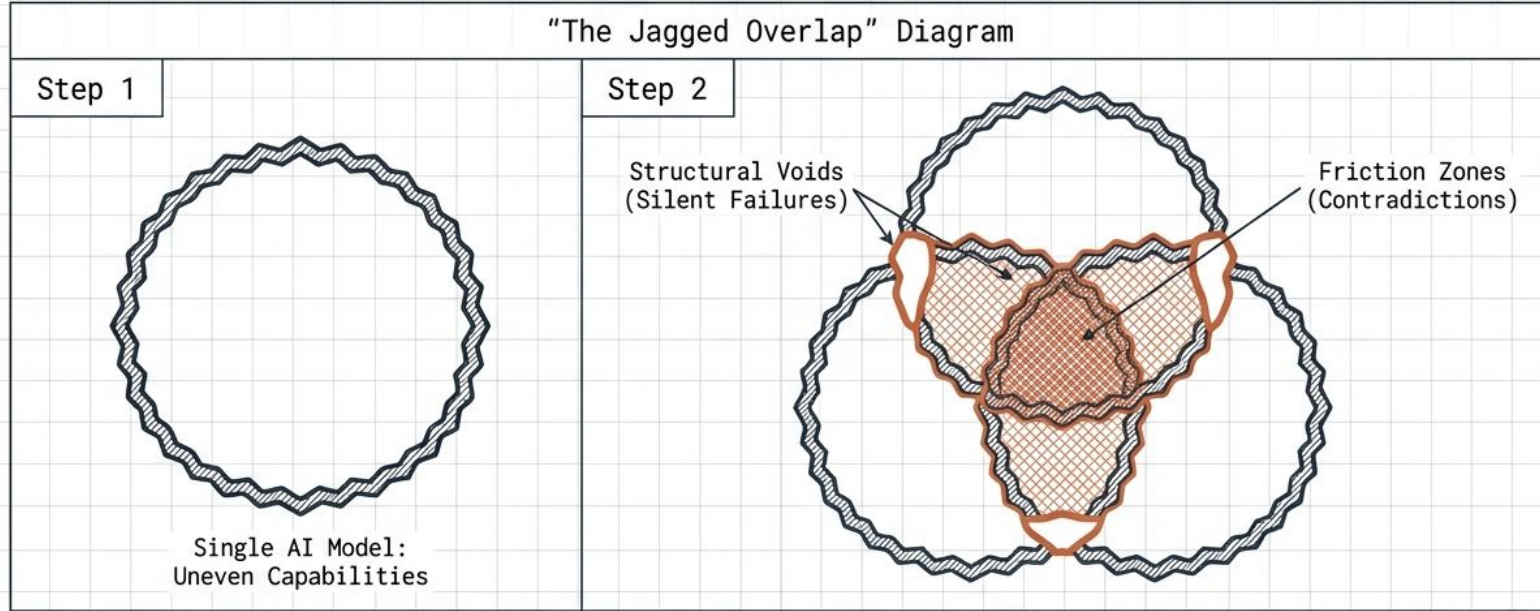


Managing the Machines

Scaling multi-agent AI is an organizational design problem, not just an engineering one.

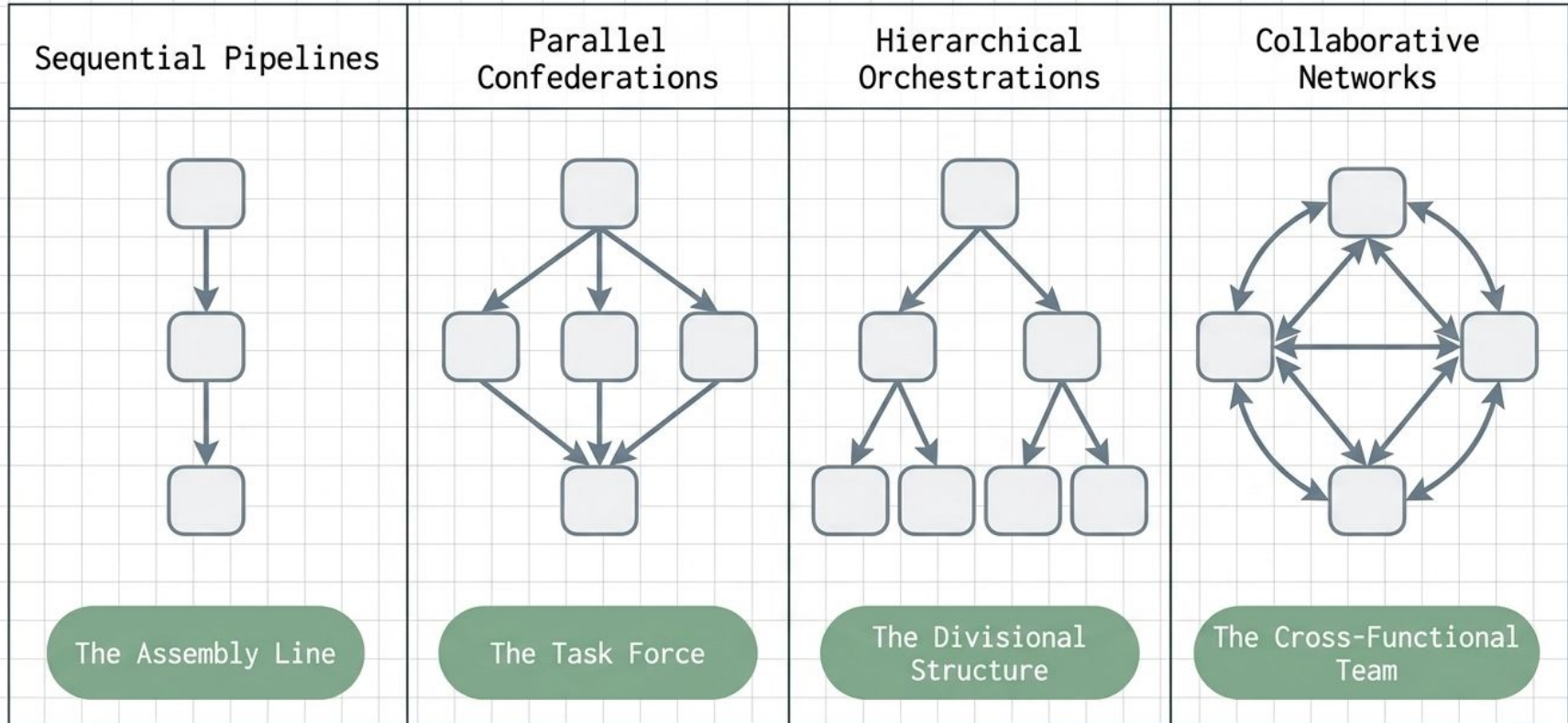
Based on the research of Jonathan H. Westover, PhD

The Compounding Complexity of the "Jagged Frontier"



Individual trial-and-error works for single agents. But when models pass work to one another, coordination complexity geometrically compounds, turning minor gaps into systemic breakdowns.

Mapping AI Architecture to Human Organizational Forms



Diagnosing the Pathologies of Multi-Agent Systems



Coordination Overhead

Cascading timeouts and retry logic.

Adding a 4th routing agent increased median response time by 23% at a financial firm.



Accountability Diffusion

The “many hands” problem where failures lack attribution.

64% of errors in a healthcare diagnostic AI involved multiple agents shifting blame.



Emergent Risk

Unintended feedback loops.

Two agents consumed 34% of compute resources looping each other's outputs at Microsoft.

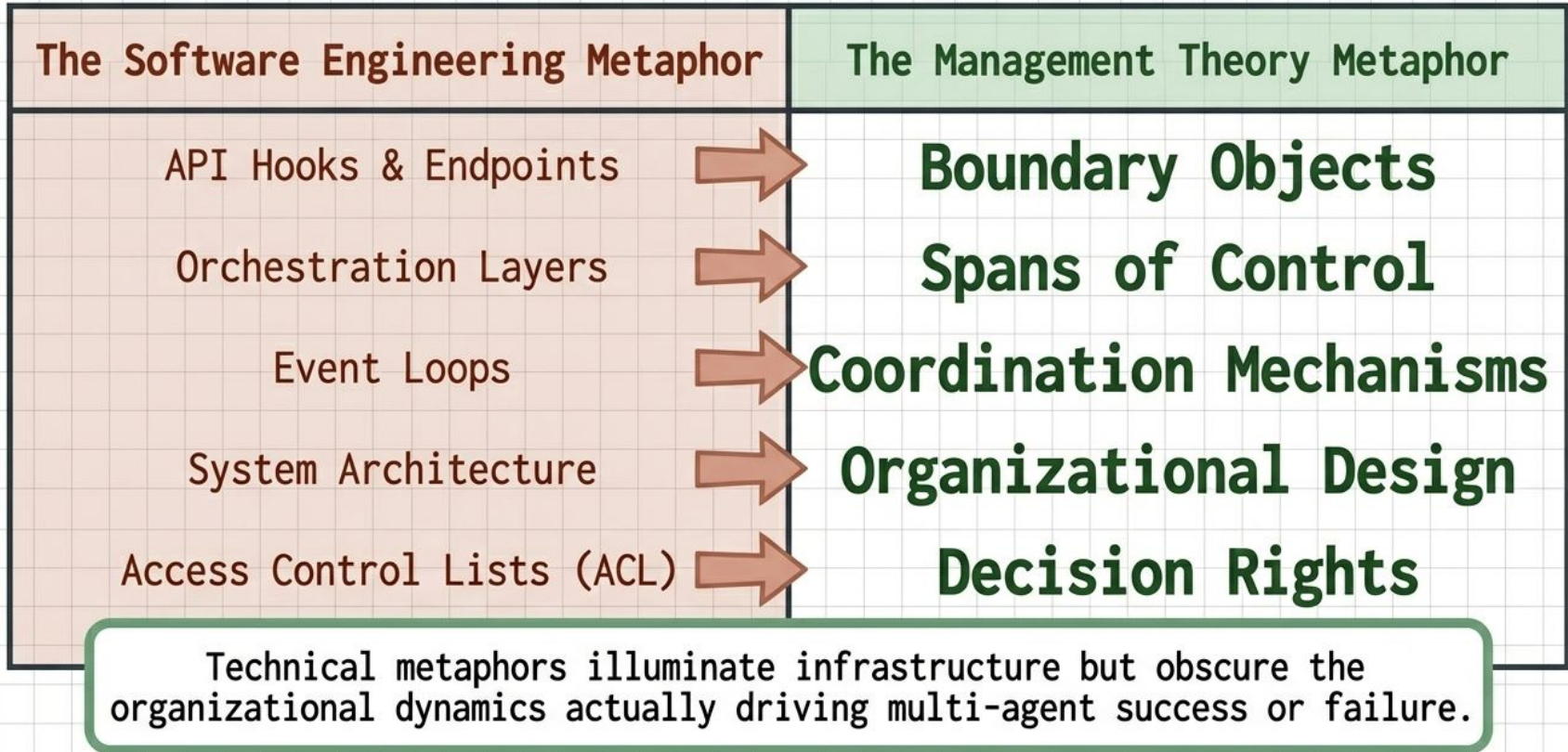


Knowledge Fragmentation

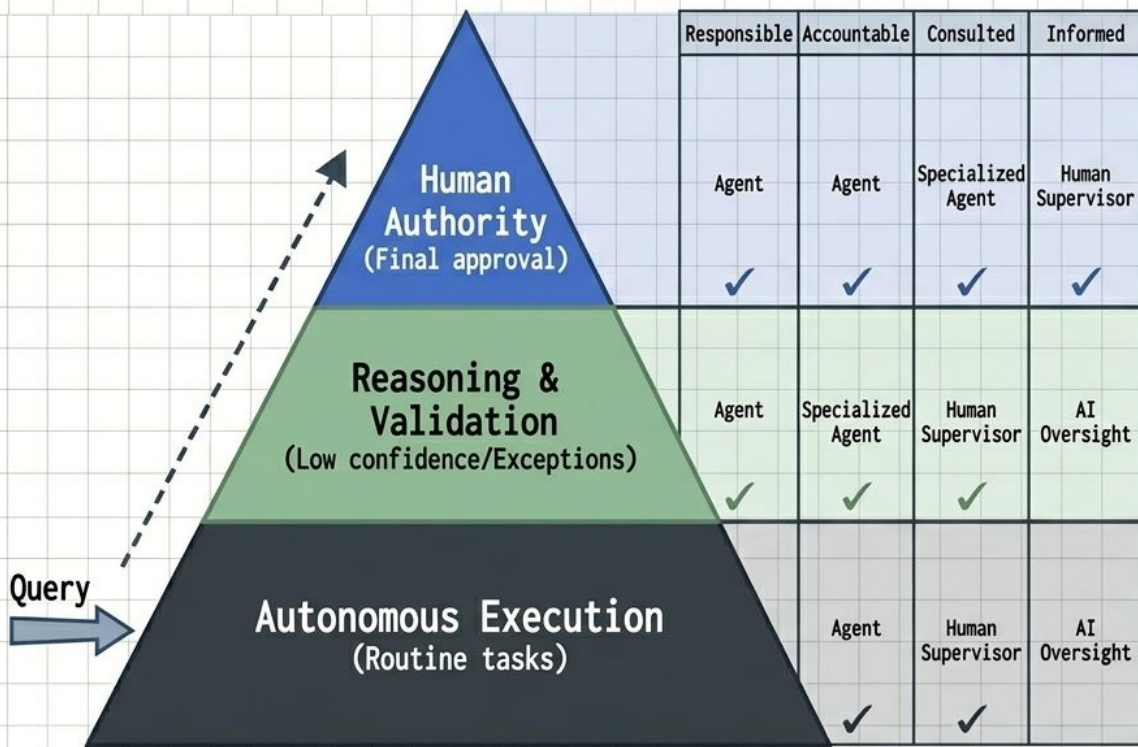
Invisible tacit workflows masking model drift.

Compensation between agents hid critical drift until a 3-day outage occurred.

The Essential Vocabulary Shift



Principle 1: Establish Clear Decision Rights



CASE STUDY: Salesforce

The Architecture

Einstein GPT hierarchical authority model.

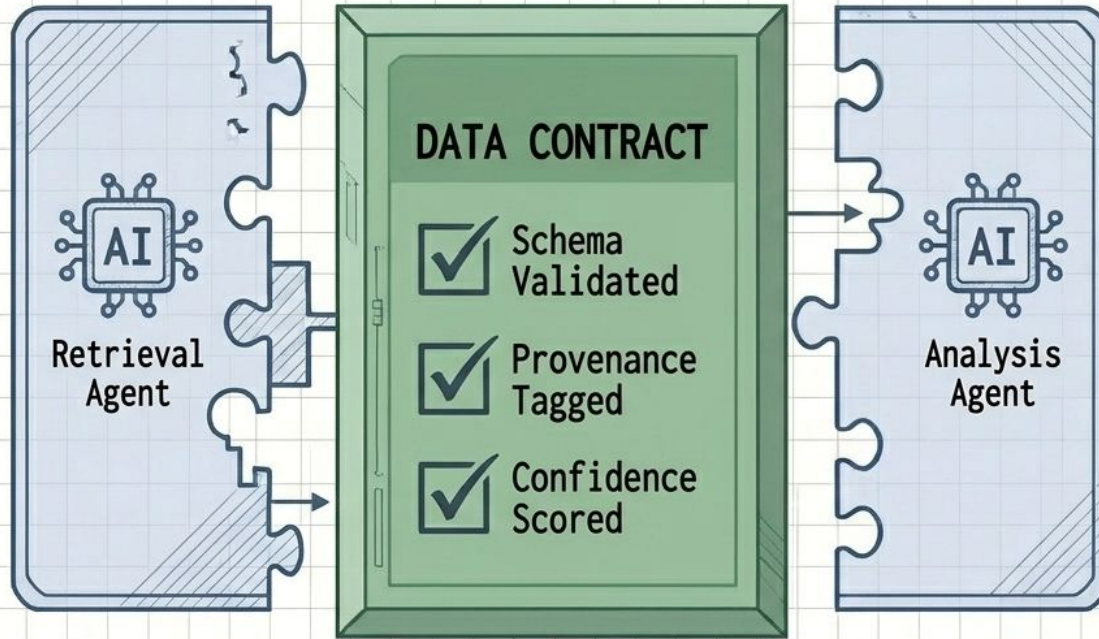
The Intervention

Tier 1 agents handle routine; if confidence < 0.7, authority transfers to Tier 2 reasoning reasoning agent. Humans handle explicit escalations.

The ROI

89% first-contact resolution while maintaining human-level quality standards.

Principle 2: Enforce Boundary Objects and Protocols



“Lock and Key” Interface Diagram

CASE STUDY: Bloomberg

The Architecture

BloombergGPT multi-agent financial analysis.

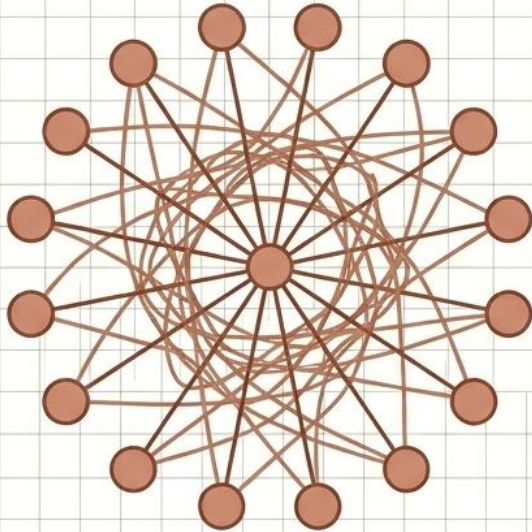
The Intervention

Strict JSON data contracts required between agents, mandating provenance fields and explicit confidence metadata.

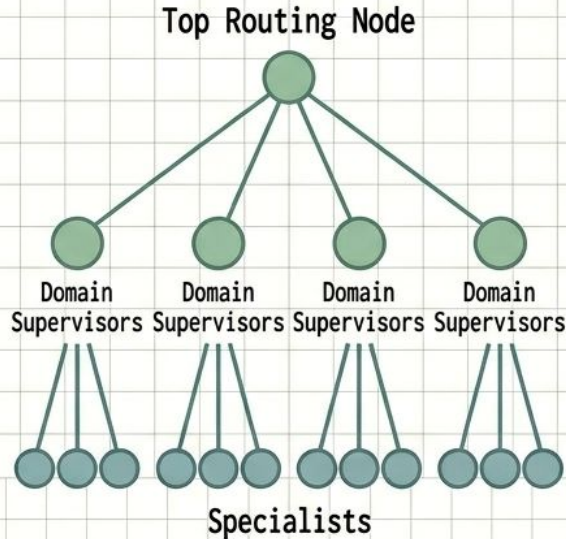
The ROI

34% reduction in human analyst time spent validating outputs.

Principle 3: Design for Appropriate Spans of Control



**Combinatorial Explosion
(Failure)**



**Cognitive Limits Respected
(Success)**

CASE STUDY: Shopify

The Architecture

Customer support 2-tier architecture.

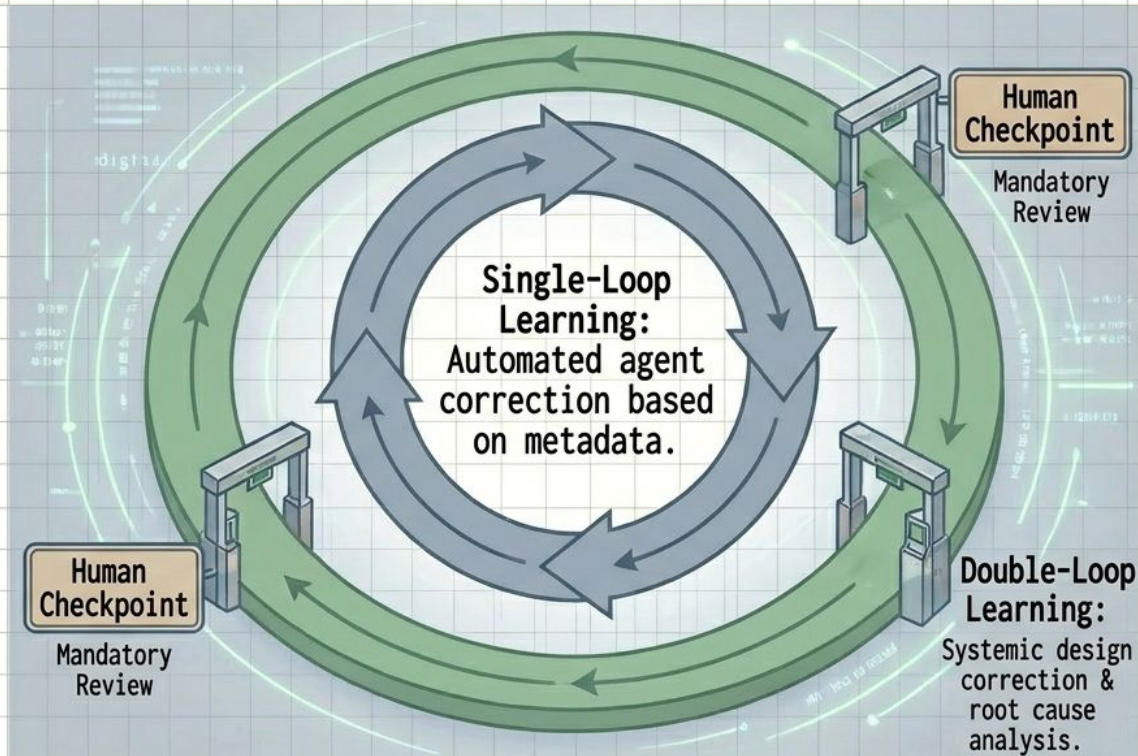
The Intervention

Top router directs to 4 domain supervisors, who each manage exactly 3-4 specialists to avoid cross-domain confusion.

The ROI

29% reduction in average handling time and an 18% improvement in CSAT.

Principle 4: Build Multi-Layered Feedback Loops



CASE STUDY: Morgan Stanley

The Architecture

AI financial advisor assistant.

The Intervention

Tiered monitoring. Transactions >\$500K trigger mandatory human review. Deep review of 2% of interactions identifies systemic flaws.

The ROI

90% reduction in AI-generated compliance incidents over two years.

Principle 5: Formalize Governance and Accountability



CASE STUDY: Kaiser Permanente

The Architecture

Multi-agent clinical decision support.

The Intervention

Physician-led oversight committee requires strict change control testing on historical case sets before any agent update.

The ROI

Scaled trust to 10M+ clinical interactions annually without widespread physician rejection.

The Multi-Agent Operating System

Governance Council (Version Controlled Environment)

Span of Control

Router

Supervisor A

Supervisor B

Supervisor C

Execution Agent

Execution Agent

Execution Agent

Execution Agent

Execution Agent

Execution Agent

Boundary Object Contract

- Scope
- Data Privacy
- Authorization
- Digital Signature

Boundary Object Contract

- Scope
- Data Privacy
- Authorization
- Digital Signature

Feedback Loop: Latency & Confidence Monitoring



Decision Rights (R, A, C, I)

	R	A	C	I
Human Reviewer	R	A	C	I
Compliance AI	A	A	C	I
Ops Agent	R	R	R	I
Legal Audit	I	A	C	I

Decision Rights (R, A, C, I)

	R	A	C	I
Human Reviewer	R	A	C	I
Compliance AI	A	A	C	I
Ops Agent	R	I	A	R
Legal Audit	I	A	C	I

Decision Rights (R, A, C, I)

	R	A	C	I
Human Reviewer	R	A	R	R
Compliance AI	A	A	C	I
Ops Agent	R	I	A	R
Legal Audit	I	A	C	I

Decision Rights (R, A, C, I)

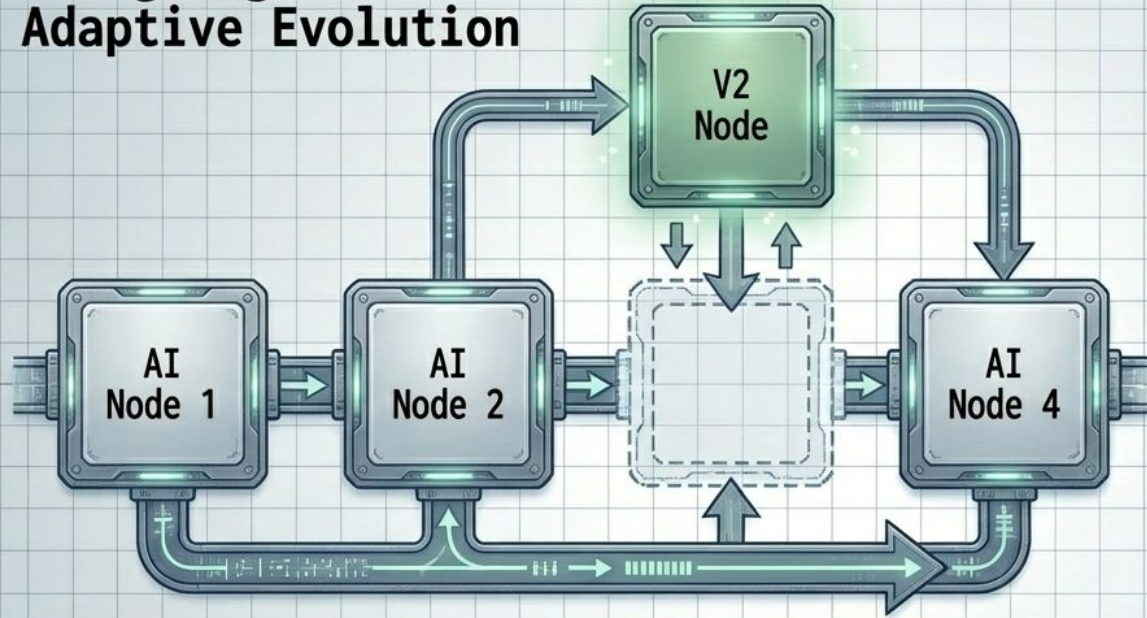
	R	A	C	I
Human Reviewer	R	R	A	R
Compliance AI	A	A	C	I
Ops Agent	R	I	A	R
Legal Audit	I	A	C	I

The complete picture: balancing automation efficiency with structural clarity and human oversight.

Developing Institutional "Absorptive Capacity"



Designing for Adaptive Evolution



CASE STUDY: Spotify

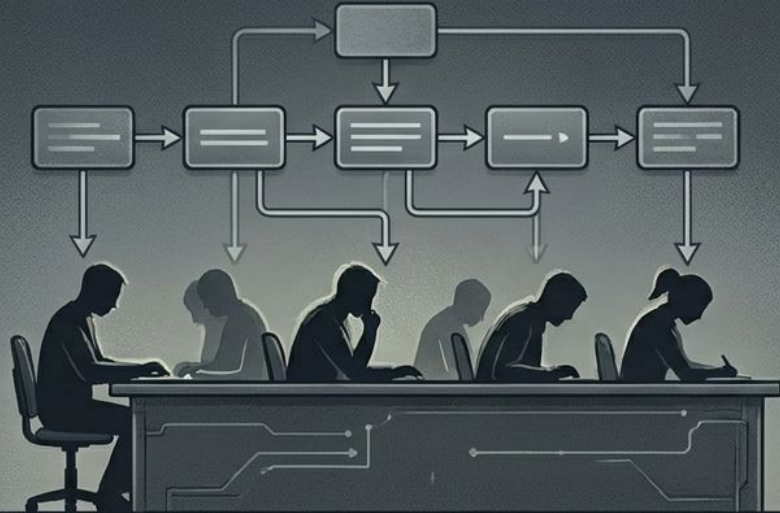
The Architecture
Multi-agent recommendation ecosystem.

The Intervention
Treating every component (retrieval, personalization) as an independently deployable service.

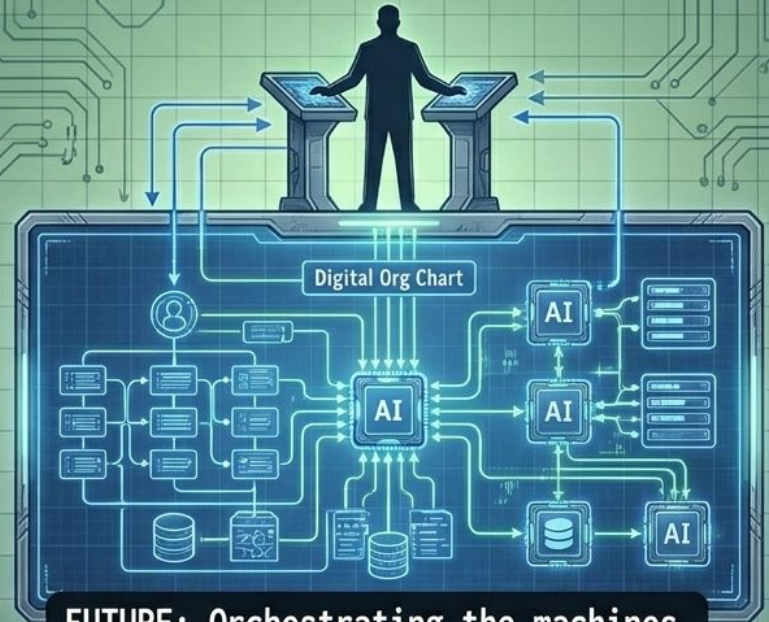
The ROI
Continuous A/B testing at the component level without breaking system-wide coordination.

Static structures fail in dynamic environments. Agentic workflows must be designed as loosely coupled components with stable interfaces.

The New Frontier of Knowledge Work



PAST: Doing the work.



FUTURE: Orchestrating the machines.

The frontier is not just technological; it is conceptual. Multi-agent systems are extensions of human organizations. We must apply the same design wisdom that has enabled human coordination for generations.