

# The Multi-Agent Scaling Wall: Why Organizational Theory is the Key to Reliable AI

## The Scaling Wall (The Coordination Crisis)

### The 35% Reliability Trap

# 35%

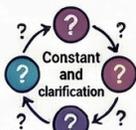


Without coordination, a 10-agent workflow with 90% individual agent reliability drops to 35% overall.



## 3x to 5x Higher Token Costs

Poorly coordinated agents waste tokens on constant clarification, reconciliation, and restating context.



## Context Fragmentation (The 'Telephone Game')

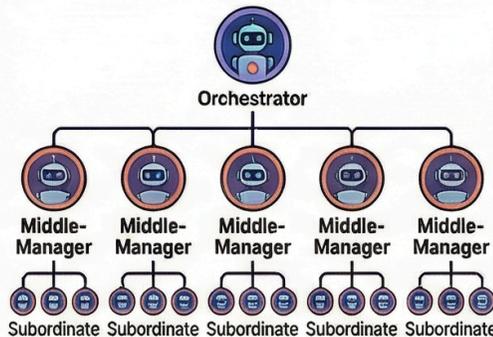


Information degrades or is misinterpreted as it passes through multiple unstructured agent handoffs.

	Ad Hoc Multi-Agent Systems	Organized Multi-Agent Systems
<b>Taken Usage</b>	3x - 5x higher (Wasteful)	Baseline (Efficient)
<b>Scaling Limit</b>	Degrades at 10+ Agents	Scalable via Hierarchies
<b>Communication</b>	Unstructured "Telephone Game"	Structured "Boundary Objects"

## The Organizational Blueprint for AI

### Apply 'Span of Control'



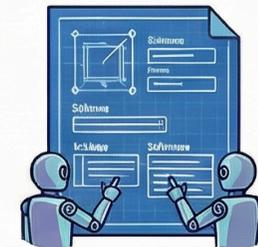
Limit orchestrators to 5-10 subordinates by introducing intermediate 'middle-manager' agent layers.

### Shift to Structured Boundary Objects

#### Freeform Text Handoff



#### Structured Boundary Object



Replace freeform text handoffs with defined schemas that multiple agents can read and write.

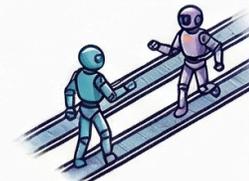
### Calibrate Coupling Mechanisms

#### Tight Coupling



Interdependent Tasks

#### Loose Coupling



Parallel, Autonomous Work

Use tight coupling for interdependent tasks and loose coupling for parallel, autonomous work.