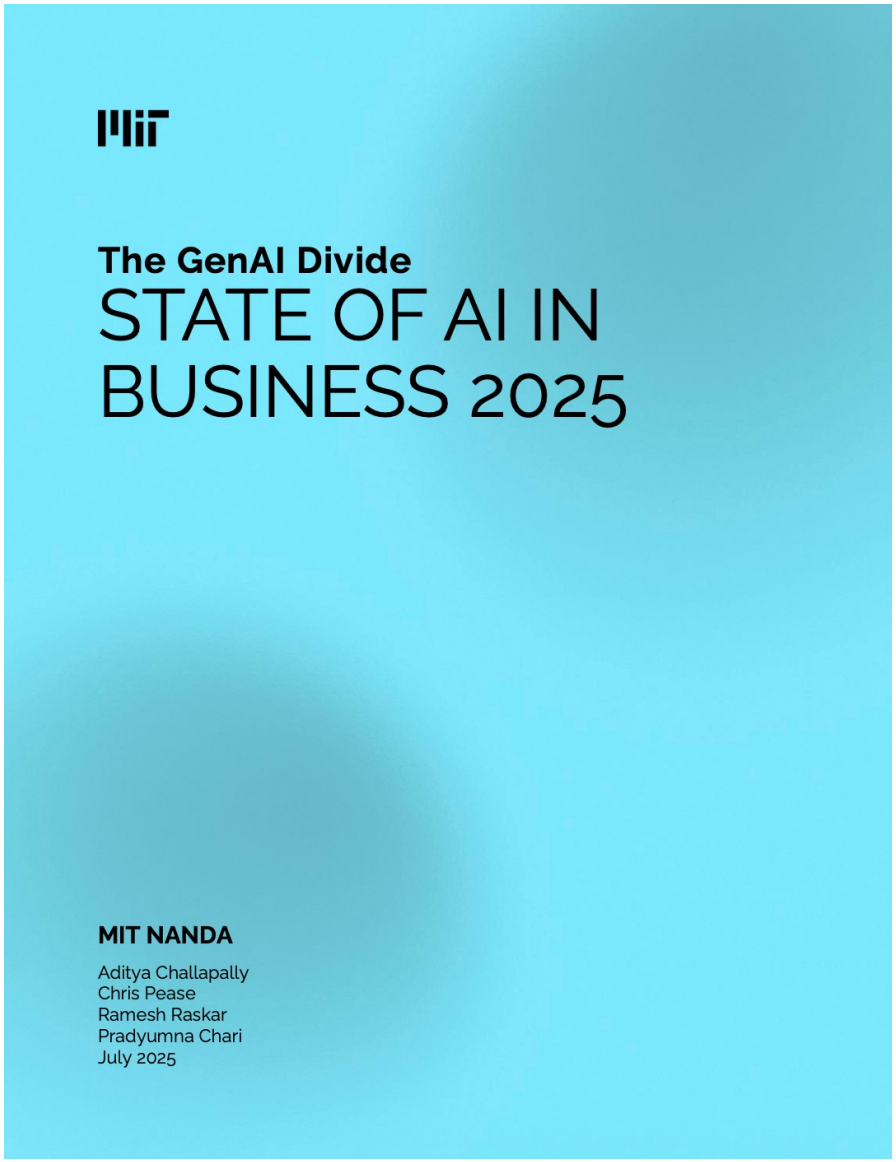
Scaling Enterprise AI Through Systems Thinking

Leader's Guide to Building a Learning Organization

:: @MashZ or @MashZahid on the socials ::



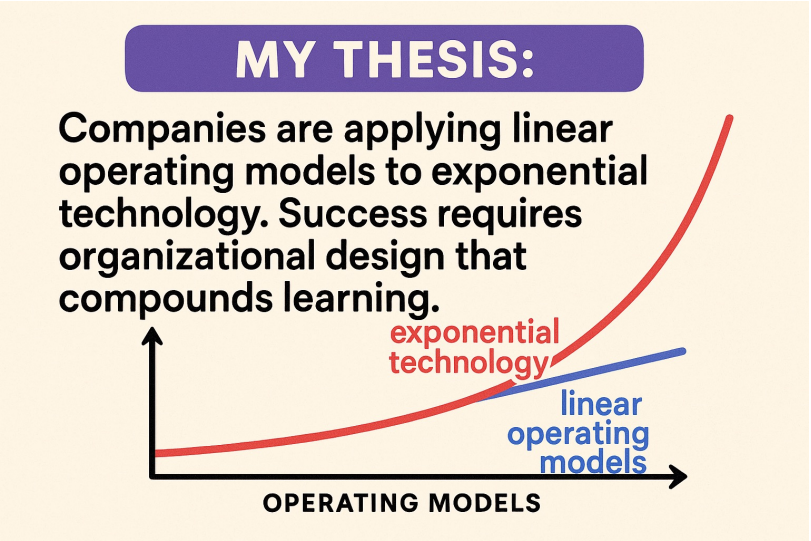
Google = MIT state of ai in business 2025

THE AI TRANSFORMATION PARADOX

MIT Study - AI makes building easier, yet failure rates are higher

What's Broken:

- Pilot-to-production chasm: ~95% of enterprise AI initiatives deliver zero P&L impact
- Learning gap: tools don't retain feedback, adapt to context, or integrate into workflows
- Misallocated spend: 50–70% of budgets in Sales/Marketing while back-office automation yields clearer ROI
- Build-bias: internal-only efforts succeed ~33% vs ~67% with external partnerships
- Centralized ownership slows scale; front-line–driven adoption with accountability performs better



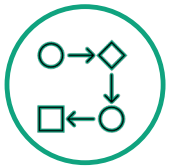
Sources: MIT NANDA State of AI in Business 2025 - https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf

THE INTEGRATION IMPERATIVE

Go beyond just building tools.

Level 1: Point Solutions
(70% of companies)

Features like Glean, Moveworks



Level 2: Process Automation
(24% of companies)

Workflow optimization, efficiency gains

Level 3: Systems Integration
(5% of companies)

Feedback loops, cross-functional orchestration

Level 4: Learning Organization
(1% of companies)

Continuous improvement, compound intelligence

Objective

I’ll show you the pattern that separates the 5% who succeed.

MASH ZAHID

Management consultant and AI engineer with real enterprise deployments plus startup experience

Roles

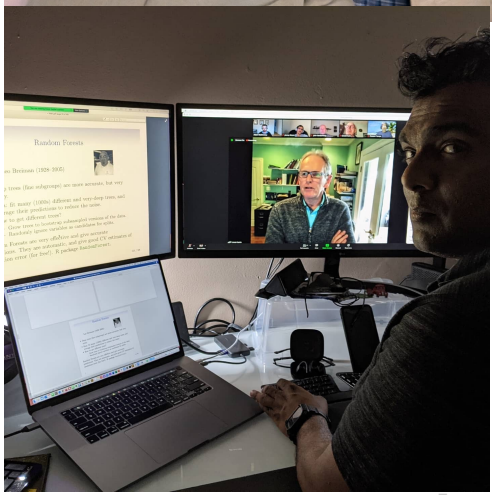
- Helping caretaker government of Bangladesh with AI Act, youth employment, and stolen-assets repatriation.
- At GM now for its executive charge for transformation using AI agents for accelerating key operations functions. Previously at IBM on engagements like conversation intelligence around customer revenue protection and collections, plus for the leading Energy and Utilities clients advising modernization of data platforms and AI automation engagements across multiple enterprise functions and enabling AI automation on SAP for financials.
- At KPMG as global director of enterprise architecture, focused on AI-led automations in Audit, Tax and Advisory lines.
- Senior strategy manager at The Home Depot, where my data-driven approach for improving retail store operations and labor strategy was detailed in a *Harvard Business Review* cover story.

Projects

- Amgen for precision medicine, cardiovascular disease Identification, and patient medication adherence
- Amazon Pharmacy, designing its growth strategy and customer expansion design, including AI powered care models
- Engine assembly automation & inventory optimization with AI methods at Subaru
- Cross-border M&A for Alghanim (Kuwaiti conglomerate), and structuring regionally relevant bond/lease instruments

Education

- The University of Chicago Booth School of Business MBA, Analytic Finance & Accounting + some Behavioral Econ PhD
- College of Wooster BA, Mathematics & Political Science



SECTION I

PATTERN RECOGNITION

1. First Principles Architecture
2. Systems Thinking
3. Talent Density
4. AI’s Multiplier Effect

1 FIRST PRINCIPLES

Building From Fundamental Truths

ANALOG THINKING

"How others do it"

Traditional Banking:

"Banks exist to ensure trust"

↓

Build a better bank

↓

1,000+ people, 14-day settlement

Method:
1. Identify the core problem (not the accepted solution)
2. Break it down to fundamental physics/math/logic
3. Rebuild from these truths up

FIRST PRINCIPLES THINKING

"Why it must be done"

Coinbase Approach:

"Trust requires verification"

↓

Build cryptographic proof

↓

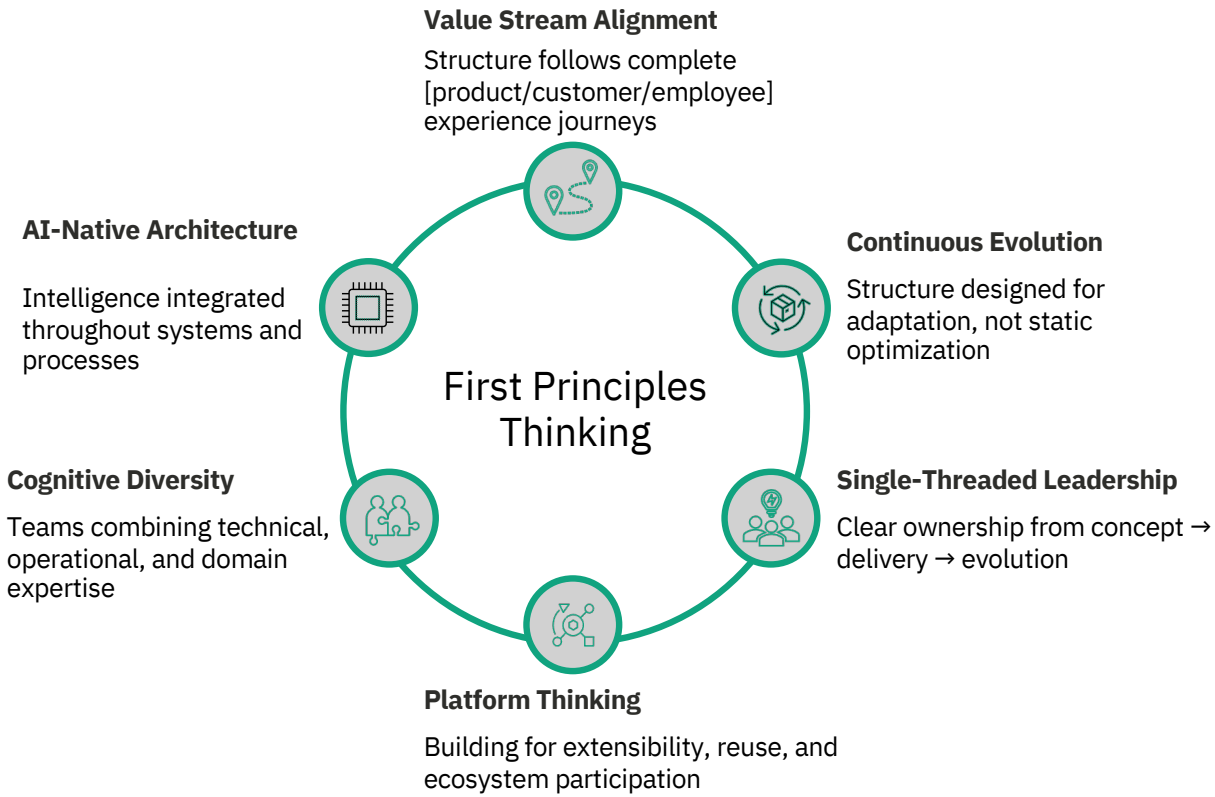
50 people, instant settlement

Assumption: "We need more people to handle more customers"
First Principle: "Customers need accurate, timely responses"
Reconstruction: "AI agents provide 24/7 accurate responses"

Result: 3 people + AI > 30 people without AI

1 FIRST PRINCIPLES THINKING

Building Systems That Endure, Evolve, and Empower For Adaptive, Intelligent Organizations



2 SYSTEMS THINKING PRIMER

“in the modern economy every organization is part of a network… and changes in one area can have side effects in others”

🔄 Systems Thinking

Holistic innovation for sustainable impact

- Recognizes and embraces complexity
- Instead of isolating one aspect, it considers the entire network of interactions.

Solutions aim to make *the whole system* more resilient and avoid the *unintended side effects*.

- It’s slower and more challenging, but yields more creative, stable solutions in the long run.

🚀 Breakthrough Thinking

“10×” innovation through bold moves
Characterized by a “move fast and break things” mentality

🎨 Design Thinking

User-centered innovation via empathy, popularized by IDEO
Focuses obsessively on the end-user’s needs and experience

“Why You Need Systems Thinking Now,” Tima Bansal and Julian Birkinshaw, in Harvard Business Review, September-October 2025, HBR Reprint R2505M

2 WHY SYSTEMS THINKING WINS IN AI TRANSFORMATION

Compound Growth vs. Additive Execution

Linear Thinking, like a waterfall **(**the 95%**):**
Breakthrough → Design → Build → Deploy → Maintain

Defect: Each stage is isolated, no learning loops

Systems Thinking, in interconnected loops (the 5% success stories)

Three Reinforcing Loops: (the "engine" of compounding growth):

1. **Learning Loop**: *Every interaction → new data → insight → system update*
2. **Talent Loop**: *Great tools → attract great talent → produce innovations → better tools*
3. **Value Loop**: *Better outcomes → more adoption → more data → better outcomes*

3 TALENT DENSITY

The Netflix Resolution

Talent Density is the concentration of exceptional talent within an organization. It reflects the ratio of high-performing individuals to total headcount—and predicts the organization's ability to scale impact per person.

Netflix vs. Traditional Media (2024)

Company	Employees	Revenue	Revenue/Employee
Netflix	13,000	$33.7B	$2.6M
Disney	225,000	$88.9B	$395K

Netflix generates **6.5x more revenue per employee** than Disney—by optimizing for Talent Density.

Source: Netflix Culture Deck, 2024 Annual Reports

Strategic Implication
Design for Talent Density → unlock exponential leverage

Now imagine amplifying this with AI agents:
Exceptional humans + intelligent automation = scalable, adaptive organizations!

4 AI'S MULTIPLIER EFFECT

The AI as Talent Multiplier Thesis - Quantified Evidence

Function	Measured AI-Amplified Improvement	Source
Customer Service	14% ↑ resolution rate	Aisera 2024
Software Development	38% ↑ performance	HBS study with 758 BCG consultants
Knowledge Work	24-38% ↑ productivity	McKinsey 2024
Code Generation	21% ↑ completion speed	Google internal study

The Compound Effect (Projected):

Year 1: Individual gains **14-38%**
Year 2: Team orchestration **2-3x estimated**
Year 3: Organizational learning **5-10x projected**

Caveat: Years 2-3 estimated are my projections based on system dynamics experience

Sources for Productivity Metrics:

1. Customer Service: 14% increase - From Aisera's LLM Agent Benchmark on Real-World Enterprise Tasks showing 14% increase in issues resolved per hour
2. Software Development: 38% performance increase - From BCG's study "GenAI Doesn't Just Increase Productivity" of 758 consultants showing 38% productivity gains
3. General Knowledge Work: 24-38% improvements - From McKinsey's "Economic Potential of Generative AI"

SYSTEMS THINKING EQUATION

Transformation equation with heuristic metrics

Traditional Approach:
Linear Growth = More People + More Features + More Capital
Result: 18-month runway → 80% failure

Exponential Growth = (Talent Density × AI Leverage) ^ Learning Velocity

SECTION II

THE FRAMEWORK

AI-POWERED SYSTEMS FRAMEWORK – 3 Pathways

- Pathway 1: Architectural Simplicity
- Pathway 2: Productive Tension
- Pathway 3: Learning Velocity

Then integrate these Pathways into a Learning Organization.

AI-POWERED SYSTEMS FRAMEWORK

Three Pathways in this AI-Powered Systems Framework

1. ARCHITECTURAL SIMPLICITY (Technical Foundation)

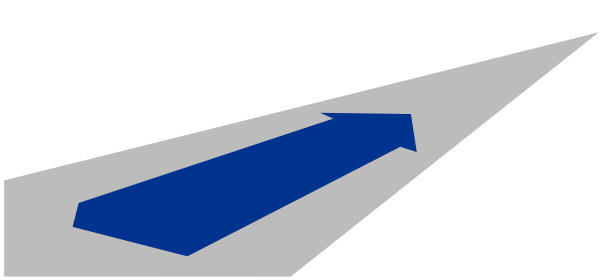
- Complexity Hierarchy: Simple (80%) → Stateful (15%) → Agentic (5%)
- "Don't agent everything" – a common-sense principle

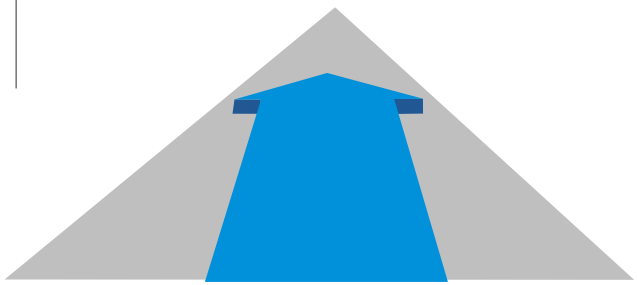
2. PRODUCTIVE TENSIONS (Organizational Dynamics)

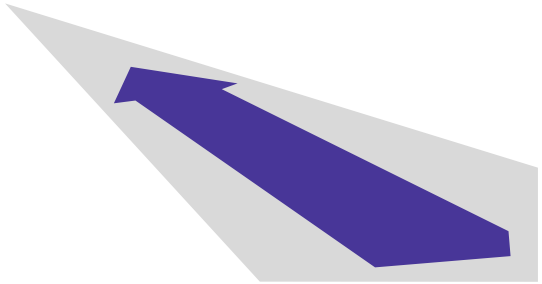
- Business vs. Technical: Speed vs. Robustness
- Human vs. AI: Experience vs. Data
- Central vs. Distributed: Standards vs. Innovation

3. LEARNING VELOCITY (Competitive Moat)

- Feedback Loop Design: Reinforcing + Balancing
- Preserve First Principles: Build & Buy
- Continuous Architecture: 8% monthly improvement







PATHWAY 1 - ARCHITECTURAL SIMPLICITY IN PRACTICE

Match Complexity To Value Creation, Not Technical Ambition

The #1 Mistake: ***Over-Engineering***

Case Study: Customer Service

Wrong AI Approach

- Multi-agent system (5 agents)
- $0.50/query, 2s latency
- 3,000 lines of code

➢ Outcome: Ran out of money in 6 months

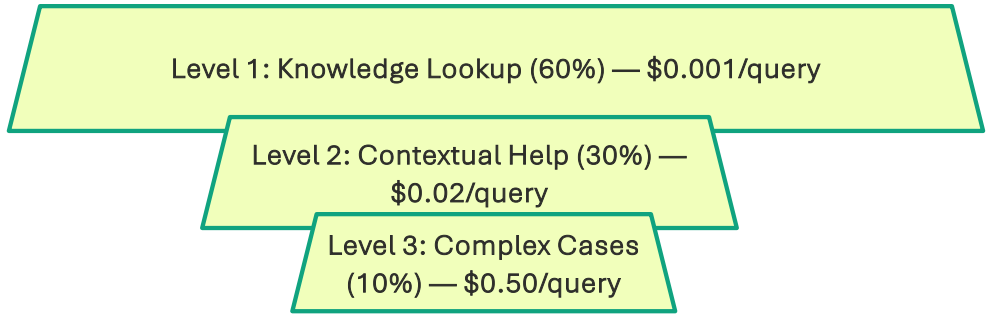
Right Approach (Tiered Model)

- Level 1: FAQ lookup - Simple RAG ($0.001/query) → 60% of queries
- Level 2: Contextual help - Stateful agent ($0.02/query) → 30% of queries
- Level 3: Complex cases – Multi-agent orchestration ($0.50/query) → 10% of queries

➢ Outcome: 85% cost reduction, >2x faster responses

Principle - Complexity should match value creation, not technical capability

+ Simplicity compounds efficiency and scalability.



PATHWAY 2 - PRODUCTIVE TENSIONS

Productive Tensions: Innovation Through Managed Conflict

Siemens Case Study

- Quality team demanded perfection <> IT product team pushed for immediate deployment
- Without choosing one, they maintained the tension.
- **Friction forced them to invent a new inspection algorithm in 10 days achieving 99.9% accuracy**

Netflix Case Study

Freedom vs. Responsibility Tension

- Engineers: “We need complete autonomy”
- Leadership: “We need aligned outcomes”
- Resolution: “Context, not control” framework
- **Result: 5x faster feature deployment, 73% fewer production issues**

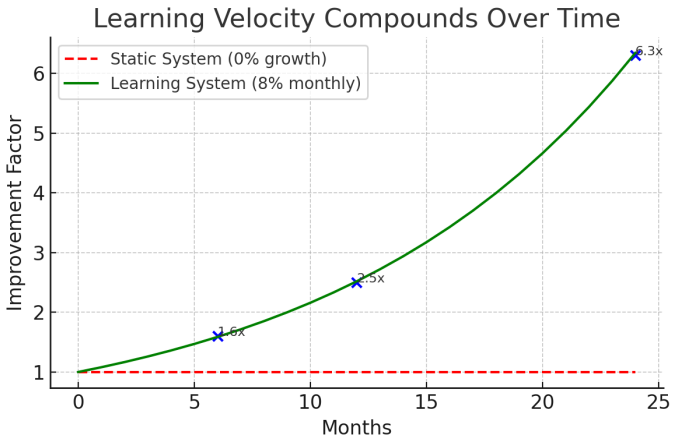
Coinbase Implementation

Decentralization vs. Compliance Tension

- Crypto ethos: “Trustless, permissionless”
- Regulatory reality: KYC & AML mandatory
- Resolution: Automated compliance via smart contracts
- **Result: 99.9% compliance rate with 90% fewer Compliance staff**

PATHWAY 3 - LEARNING VELOCITY

Compound Advantage By Balancing Speed with First Principles Innovation



Measurement Framework:

- **Static System**: 0% monthly improvement → Decline "death spiral"
- **Learning System**: 8% monthly improvement compounds
  - 6 months → 1.6x
  - 12 months → 2.5x
  - 24 months → 6.3x

First Principles Decay Problem – When teams only apply pre-built AI solutions, innovation from fundamentals erodes.

"*We risk creating a generation of researchers comfortable with applying existing solutions rather than developing new ones.*" – 2024 O1 Replication Study

Netflix's **Hybrid Learning** Architecture:

- 70% of engineers use pre-built tools → speed
- 30% build from first principles → innovation
- *And* rotation ensures everyone builds core systems periodically

Result: 10x faster deployment **with sustained innovation**

BUILDING LEARNING ORGANIZATIONS

Peter Senge's Five Disciplines adapted for age of AI amplification

SYSTEMS THINKING (Enhanced by AI)

Traditional: Human pattern recognition

AI-Enhanced: ML reveals hidden system relationships

e.g., Netflix's algorithms discover viewing patterns humans missed

PERSONAL MASTERY (AI-Augmented)

Traditional: Individual skills development

AI-Enhanced: Personalized learning paths, skill gap analysis

e.g., Coinbase uses AI to identify optimal role transitions

MENTAL MODELS (AI-Challenged)

Traditional: Assumptions go unchallenged

AI-Enhanced: Data surfaces unconscious biases

e.g., AI revealed 67% of "technical" rejections were bias

SHARED VISION (AI-Distributed)

Traditional: Top-down vision statements
AI-Enhanced: Analyze all stakeholder input, personalize communication
Result: 3x higher vision alignment

TEAM LEARNING (AI-Accelerated)

Traditional: Meetings and retrospectives
AI-Enhanced: Real-time dialogue analysis, optimal team composition
Outcome: 40% project completion

Multiplier Effect

Each discipline reinforces others exponentially

IMPLEMENTATION FORMULA

Phase 1: **Map one process → Find the feedback loop**

Phase 2: **Build simple baseline → Add one agent**

Phase 3: **Introduce tension(s) → Measure learning rate**

Phase 4: **Scale or pivot**

- ❑ Identify natural opposing forces in your organization
- ❑ Create forums for structured disagreement (not consensu
- ❑ Measure innovation velocity at tension points
- ❑ Adjust tension levels based on output quality

SECTION III

CODE REVIEW

- Code notebook is on GitHub
- Healthcare model demo for AI systems composed of goal-driven agents

PHASE 1: MAP ONE PROCESS → FIND THE FEEDBACK LOOP

```python
# Multi-agent system for patient journey optimization
care_coordinator = Agent(
    name="Care Coordinator",
    instructions="Coordinate between providers and ensure care continuity for patients"
)

benefits_navigator = Agent(
    name="Benefits Navigator",
    instructions="Optimize insurance coverage and identify cost-saving opportunities"
)

health_coach = Agent(
    name="Health Coach",
    instructions="Personalize patient engagement strategies and improve adherence"
)

clinical_analyst = Agent(
    name="Clinical Analyst",
    instructions="Identify care gaps and clinical risks using evidence-based guidelines"
)

print("✅ All healthcare agents created successfully")
print(f"Agents: {[agent.name for agent in [care_coordinator, benefits_navigator, health_coach, clinical_analyst]]}")
```

```
✅ All healthcare agents created successfully
Agents: ['Care Coordinator', 'Benefits Navigator', 'Health Coach', 'Clinical Analyst']
```

The patient deep dive showing how multiple agents (Risk, Care Gap, Benefits, Engagement) create productive tension.

This shows the natural opposing forces - clinical urgency vs. cost optimization vs. patient engagement barriers.

PHASE 2: BUILD SIMPLE BASELINE → ADD ONE AGENT

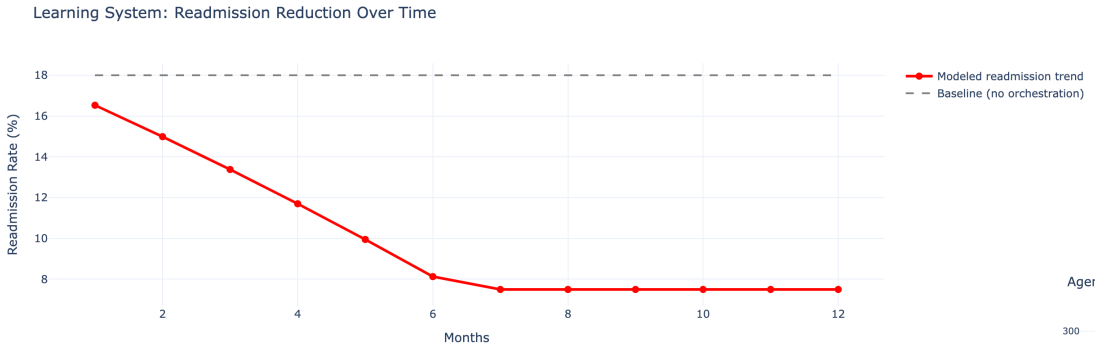
```python
# Cell 3: Demonstrate "Don't Agent Everything" Principle
def get_guideline_targets(guidelines: dict) -> str:
    diabetes = guidelines['diabetes']
    cardio = guidelines['cardiometabolic']
    return (
        f"Targets → HbA1c: <{diabetes['a1c_target']}%, BP: <{cardio['bp_systolic_target']}/{cardio['bp_diastolic_target']} mmHg, "
        f"LDL: <{cardio['ldl_target']} mg/dL"
    )

simple_answer = get_guideline_targets(synthetic_data.guidelines)
print("✅ Simple lookups stay simple — no agent orchestration required")
print(simple_answer)

print("💡 Use direct functions for:")
print("- Static clinical targets")
print("- Quick eligibility checks")
print("- Single data point retrieval")
```

```
✅ Simple lookups stay simple — no agent orchestration required
Targets → HbA1c: <7.0%, BP: <130/80 mmHg, LDL: <100 mg/dL
💡 Use direct functions for:
- Static clinical targets
- Quick eligibility checks
- Single data point retrieval
```

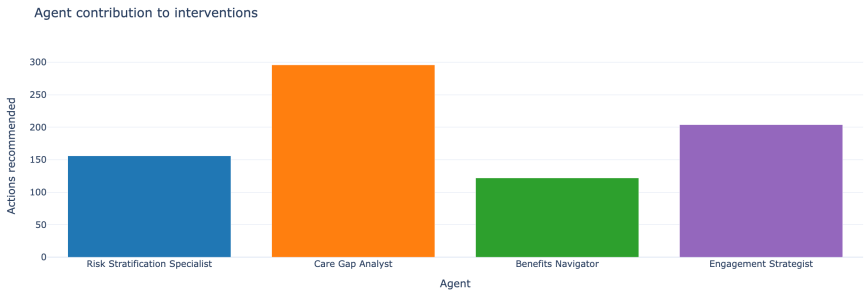
Your “Don't Agent Everything” principle!

Contrast this with the multi-agent orchestration results yielding 639 interventions across 296 gaps.

PHASE 3: INTRODUCE TENSION(S) → MEASURE LEARNING RATE



Readmission reduction from 18% to 7.5% over 12 months



Agent contribution distribution (296 actions from Care Gap Analyst vs 122 from Benefits Navigator)

Care gap resolution acceleration from 17.6% to 79.2%

PHASE 4: SCALE OR PIVOT

Impact quantification:

```
Traditional care management (30 FTE) → 18% readmission, $450 per patient/month
Our orchestrated pod (3 FTE + agents) → 7.5% readmission, $125 per patient/month
Annualized savings modeled: $182,340 (~$1,520 per patient)
Agents recommended 639 interventions across 296 gaps
Human team focuses on coordination + escalations; agents handle analysis, monitoring, and follow-up queues.
```

Now go slay!

Professions are "conspiracies against the laity."

https://en.wikipedia.org/wiki/Conspiracies_against_the_laity

Regarding such conspiracies, Tim Harford argued the following in his 2006 book The Undercover Economist, that "doctors, actuaries, accountants and lawyers manage to maintain high wages through... erecting virtual 'green belts' to make it hard for competitors to set up shop. Typical virtual green belts will include very long qualification periods and professional bodies that give their approval only to a certain number of candidates per year. Many of the organizations that are put forth to protect us from 'unqualified' professionals in fact serve to maintain the high rates of the 'qualified' to whom we are directed."