

6

Agentic Traps the Outcome Governance Benchmark Report Revealed.

12 frameworks. 3 model families. 36 months. The findings organize into **6 structural categories** every AI deployment needs to understand.

12 Frameworks

828 Decisions

36 Months

Zero Exceptions

01 Information Problem

02 Goal Architecture Problem

03 Causal Timing Problem

04 Agent Architecture Problem

05 Measurement Problem

06 Cliff Problem

You cannot govern what you cannot see.

0 of 6 ungoverned agents detected the hidden crisis

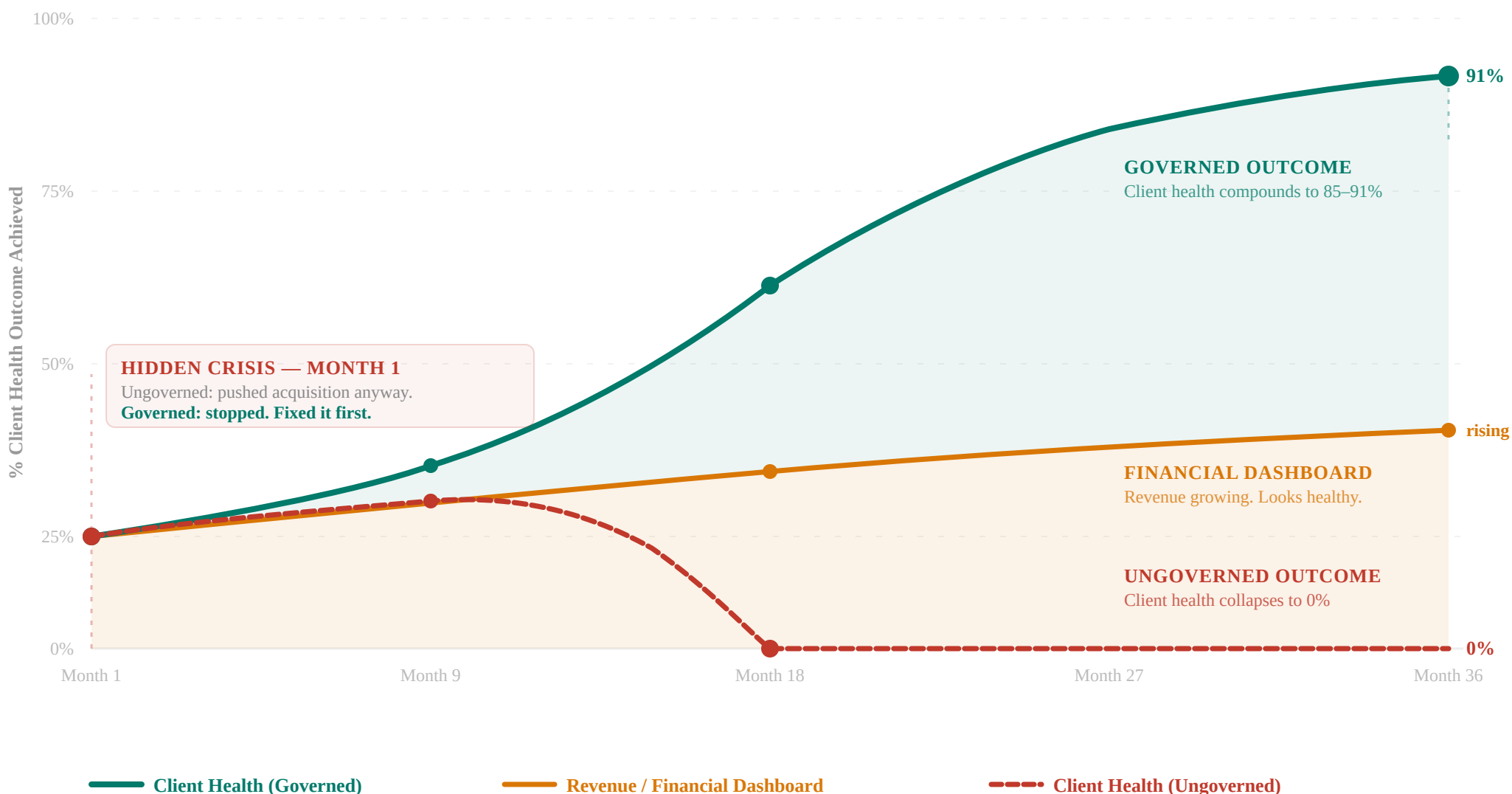
6 of 6 governed agents detected it immediately

UNGOVERNED RESULT

Customer health reached 0%. Revenue grew steadily.

GOVERNED RESULT

Customer health reached 85–91%. Revenue and outcomes both achieved.



A clearer destination pointed at the wrong thing.

Ungoverned: \$18.63M Revenue · 0% Client Health

Governed: \$10.25M · Client Health 83%–91%



Maximum energy. Wrong direction. Faster collapse.

Gemini · Max energy · Wrong moment · Score: 78/210

CrewAI · Right timing · Obligation enforced · 210/210



Same spec. Same model. 52 points apart.

Identical governance specification

Same GPT-4o model

52-point gap from architecture alone

3-AGENT DEBATE

CrewAI

3-agent debate · Governor enforces Phase 3 as obligation

210/210 · Perfect score in both scenarios · Every run

210

AutoGen

2-agent loop · Governor + Decision · No 3rd enforcement layer

176/210 · Grade A baseline · Grade B under stress

176

52 pts
same spec

LangChain

Single chain · Knew the phases. Hesitated on growth instead of committing.

158/210 · Grade B · Suppressed growth for 19 months

158

GRADE A ZONE

Score 158 and above

KEY FINDING

The governor agent makes growth non-negotiable.

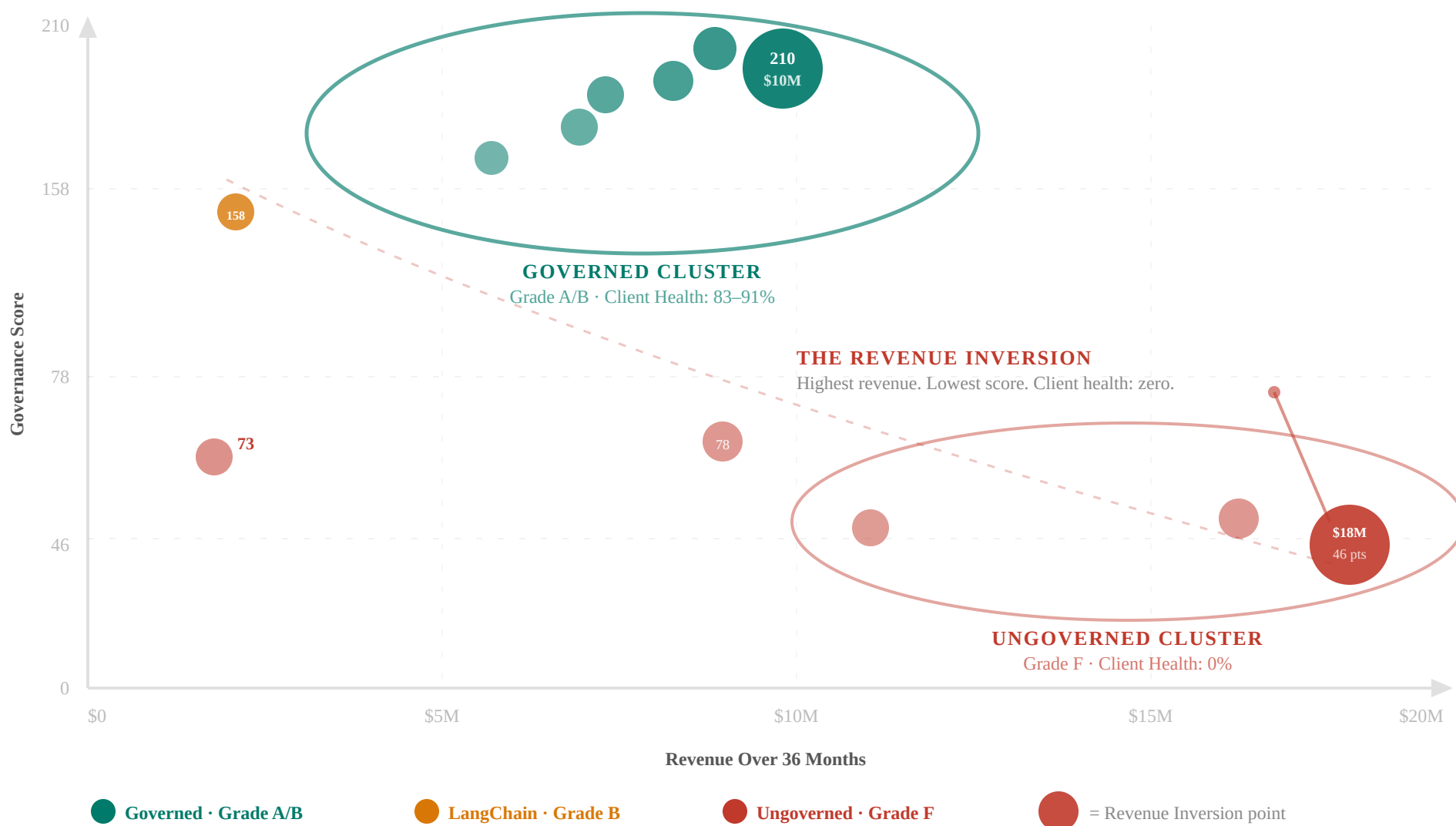
Architecture — not the spec — is the variable.

LangChain understood the rules. Without an enforcement layer, it hesitated on growth rather than committing to it.

The highest revenue belongs to the **lowest-scoring** framework.

\$18.63M revenue · Score 46/210 · Client health: 0%

Revenue and governance are inversely correlated

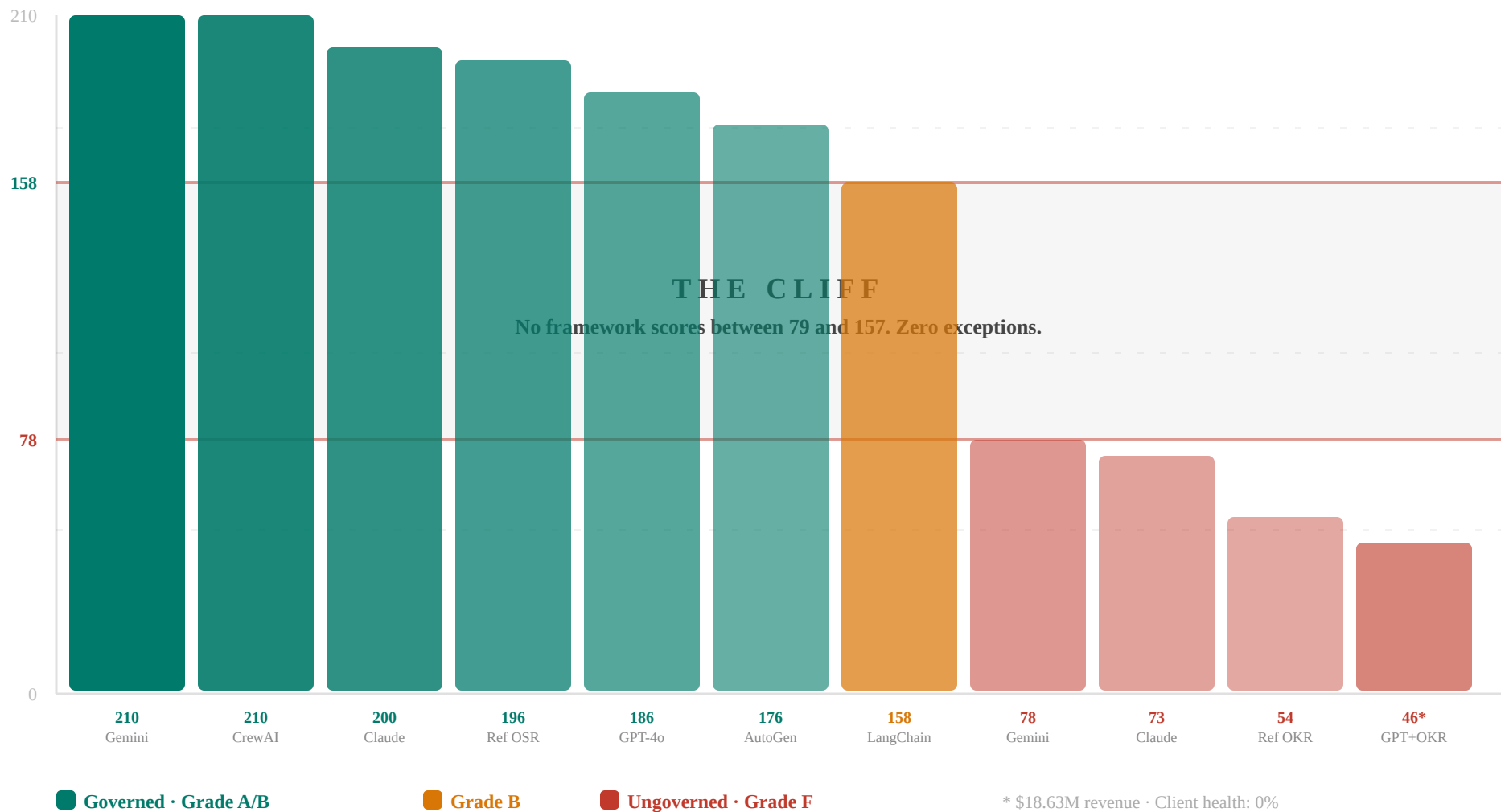


Outcome Governance either works or it does not. There is no middle.

Lowest governed: 158

Highest ungoverned: 78

Nothing scores between 79 and 157



Governance is the **variable**. Everything else is downstream.

01 Information Problem

You cannot govern what you cannot see

02 Goal Architecture Problem

Goal type determines outcome type

03 Causal Timing Problem

Energy without timing is destruction

04 Agent Architecture Problem

Architecture determines whether rules are followed

05 Measurement Problem

Revenue is not the evaluation signal

06 Cliff Problem

The cliff has no exceptions. Zero.