PORTFOLIO WRITE-UP

Autonomous Machine: simulating institutional decline and AI-enabled recovery

A parametric agent-based simulation built against my thesis: institutions can become autonomous machines, where architecture rather than conscious intent keeps filtering reality, consuming capacity, resisting correction, and producing failure patterns.

PYTHON SIMULATION | 30,500+ RUNS | REAL-WORLD UK COUNCIL + NHS DATA | AI ADOPTION DYNAMICS

Project Thesis

The title Autonomous Machine refers to an institution whose architecture has become self-propelling. The organisation still contains people making local decisions, often for understandable reasons, but its reporting layers, budget controls, committees, incentives, risk processes, and accountability boundaries start to behave like machinery. They keep processing information and resources in the same pattern even when that pattern is no longer adaptive.

The starting thesis is that institutional failure is not usually caused by one bad person, one bad policy, or one missing dashboard. Large organisations build structures for valid reasons: coordination, safety, accountability, auditability, and political control. Over time those same structures can become resistant to change. Signals get softened as they move upward, visible problems outcompete severe but delayed problems, overhead becomes easier to defend than frontline delivery, and accumulated backlog reduces the institution's ability to recover. The simulation tests whether decline can emerge from that machine-like architecture even when individual agents are not malicious.

From Thesis To Model

The simulation operationalizes that thesis by turning it into testable claims about delivery trajectories, crisis timing, backlog accumulation, and intervention effects. In that sense, the "machine" is not a metaphor for robots or automation; it is the operating logic produced by the institution's structure. The model focuses on the parts of the argument that can be represented directly: resource flow, signal filtering, agent quality, morale, overhead, performative spend, backlogs, and crisis dynamics.

MODELED MECHANICS

- **Resource flow:** money moves through leadership, management, and frontline layers before reaching services.
- **Signal filtering:** true need and backlog can be softened as information moves upward.
- **Capacity loss:** morale depletion, high-competence departures, degraded replacement pools, overhead, and performative work reduce delivery.
- **Resilience loss:** accumulated backlog and crisis costs drain future budgets and make recovery harder.
- **Four tracked losses:** accuracy, capacity, legitimacy, and resilience.

What It Is

Autonomous Machine is a research simulation for institutional resource allocation. Money enters a layered organisation, passes through leadership and management, and eventually reaches frontline service areas. Along the way it can be consumed by overhead, reporting loops, performative work, crisis response, poor signals, and capability loss.

The point was not to build a generic management toy. The simulation asks a more specific question: when an institution is under budget pressure and rising demand, which feedback loops make decline self-reinforcing, and which interventions actually restore delivered service?

The Model

Each run tracks services with true demand, visible demand, backlog, consequence delays, visibility, statutory pressure, and delivery outcomes. Signals flow upward through the hierarchy and budgets flow back down, creating room for signal filtering, management overhead, political bias, gaming, and delayed consequences.

FOUR LOSSES TRACKED

- **Accuracy:** true need becomes distorted as it moves upward.
- **Capacity:** money and staff time are absorbed before reaching delivery.
- **Legitimacy:** visible performance can drift away from lived outcomes.
- **Resilience:** accumulated backlog reduces the ability to recover later.

The core levers include demand/budget scissors, overhead creep, performative spend, brain drain, signal corruption, statutory service floors, reserves, borrowing, and crisis budget drain. These can be varied

independently, which makes it possible to test decline as an interaction between pressures rather than a single-cause story.

Validation

The council model was calibrated against UK local government data and checked against five real Section 114 events. After adding fiscal buffers, statutory floors, and smoother crisis dynamics, the current hindcast reached a median timing error of 1.8 years, with three of five cases matched within two years. I treat those hindcasts as informed calibration checks rather than a claim of full predictive validation: the model is most useful as an executable thesis about mechanisms, not as a finished forecasting product.

30,500+
simulation runs across experimental conditions

1.8 yr
median timing error in informed council failure hindcasts

3/5
Section 114 cases matched within two years in calibration checks

Figures are from the project simulation outputs and hindcast reports; they should be read as calibration evidence, not external validation.

The Replacement Nervous System

The thesis does not frame AI as generic automation. It frames it as a possible replacement nervous system for institutions: a way to route signals around career-motivated filtering, connect decisions to outcomes, and make long-term consequences visible before they become crises.

That framing shaped the simulation design. AI was represented as three architectural layers that address the specific failures in the thesis rather than as a single productivity multiplier.

INTERVENTION LAYERS

- **Sensory layer:** captures true signal alongside filtered human signal, detects performative spend, and reduces frontline morale depletion.
- **Accountability layer:** makes allocation decisions and outcomes visible, compresses unjustified overhead, and slows integrity erosion.
- **Intelligence layer:** projects backlog, dampens electoral or inspection bias, supports triage, reduces demand through early intervention, and can route some budget directly to frontline delivery.

AI Was Modeled As Adoption, Not Magic

The AI part of the model was deliberately made messy. AI adoption follows an S-curve, leadership trust lags adoption, gaming adapts to evade detection, management resistance grows when budget bypass threatens

identity or control, and implementation has a real cost.

Findings

In degraded systems, AI improved delivery efficiency substantially by restoring signal quality and reducing wasteful flow through the hierarchy. In the current median comparison, activating all three AI layers at tick 50 raised final delivery efficiency from 25.9% to 47.0%. The stricter need-met metric rose from 7.2% to 13.1%, showing the useful but limited effect of better institutional flow when accumulated backlog is still large.

> **25.9%**
> median final delivery efficiency without AI

> **47.0%**
> median final delivery efficiency with all three AI layers

> **13.1%**
> median final current need met with AI active

The important caveat is that delivery efficiency is not the same as need actually met, and AI does not erase accumulated backlog. It improves flow and reveals distortion, but institutions that have already deferred too much work still need operational repair and capital investment.

What I Built

- A Python simulation engine for layered institutional resource flow, signal distortion, service backlogs, and crisis dynamics.
- Batch experiment tooling for large parameter sweeps, sensitivity tests, hindcasts, and cross-sector comparisons.
- Separate council and NHS scenario models with calibrated service areas, demand growth, consequence delays, and backlog effects.
- An AI intervention model that includes adoption, trust lag, resistance, gaming adaptation, implementation cost, and political capture effects.
- Analysis outputs and charts for survival, phase windows, AI decomposition, money flow, visibility traps, and institutional health.

Prepared as a concise project write-up for portfolio review. Source materials include my private Autonomous Machine thesis draft, simulation code, the "Institutional Decline as Emergent Architecture" paper draft, findings summary, and generated experiment outputs.