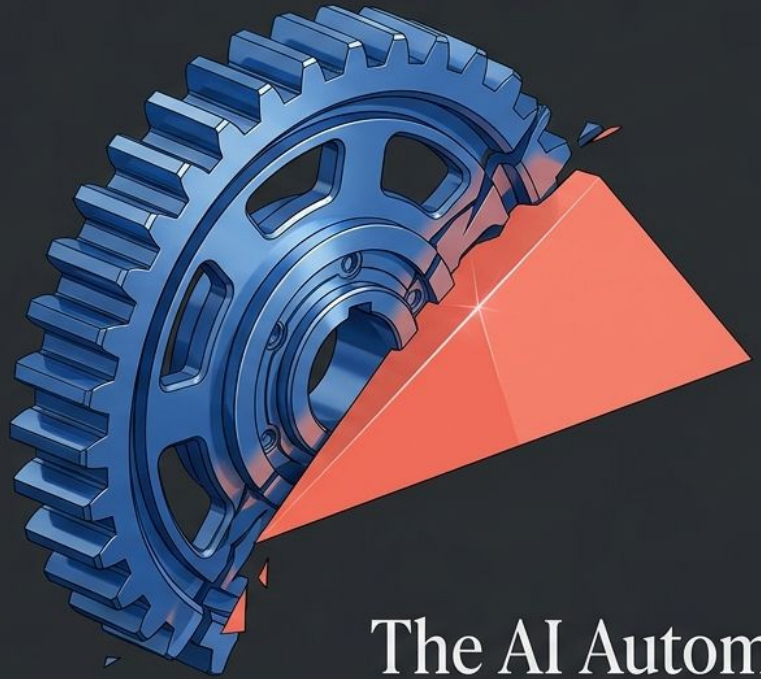


The AI Automation Paradox

Why perfect foresight cannot stop the race to the cliff.

Based on the research of Jonathan H. Westover, PhD.

The cognitive displacement wave is accelerating

100k+

Tech workers displaced in 2025.
Over **80%** of U.S. workers hold jobs with task that their tasks susceptible to Large Language Models.

50%

Block Inc. eliminated half its workforce (4,000 employees) in early 2026.

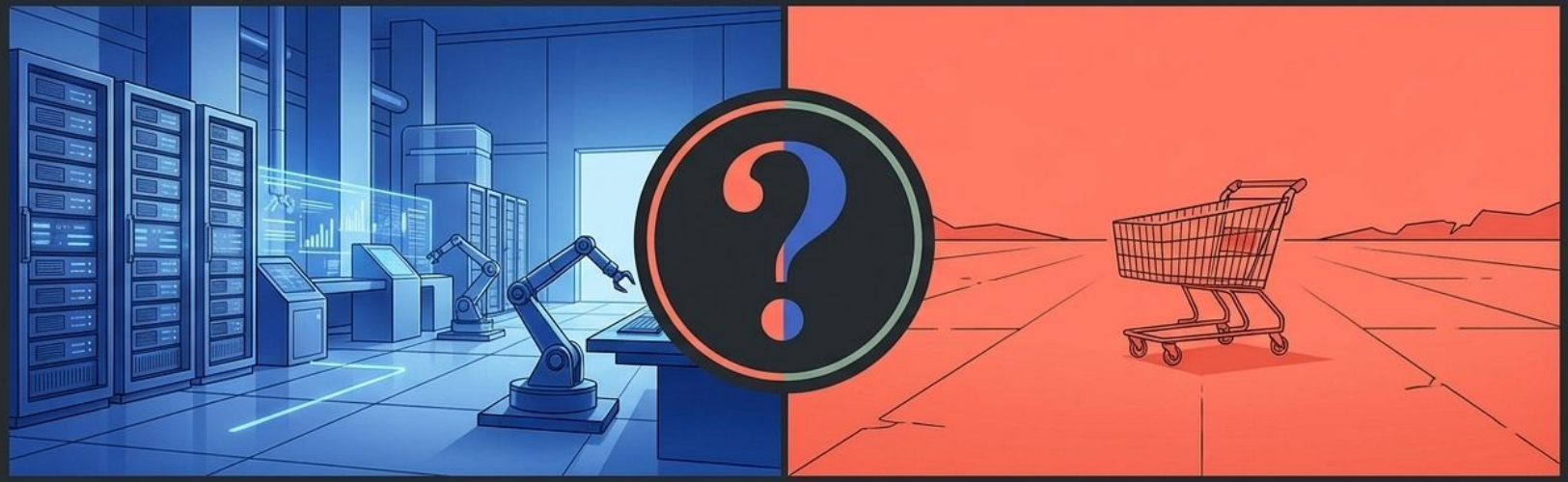
Salesforce replaced 4,000 agents with AI.

5x

Goldman Sachs and **Infosys** deployed **autonomous coders,** enabling one senior engineer to execute work previously requiring five-person teams.

"Within the next year, the majority of companies will reach the same conclusion." — *Jack Dorsey*, CEO of Block Inc. (Feb 2026)

The core paradox: Dead consumers buy no products



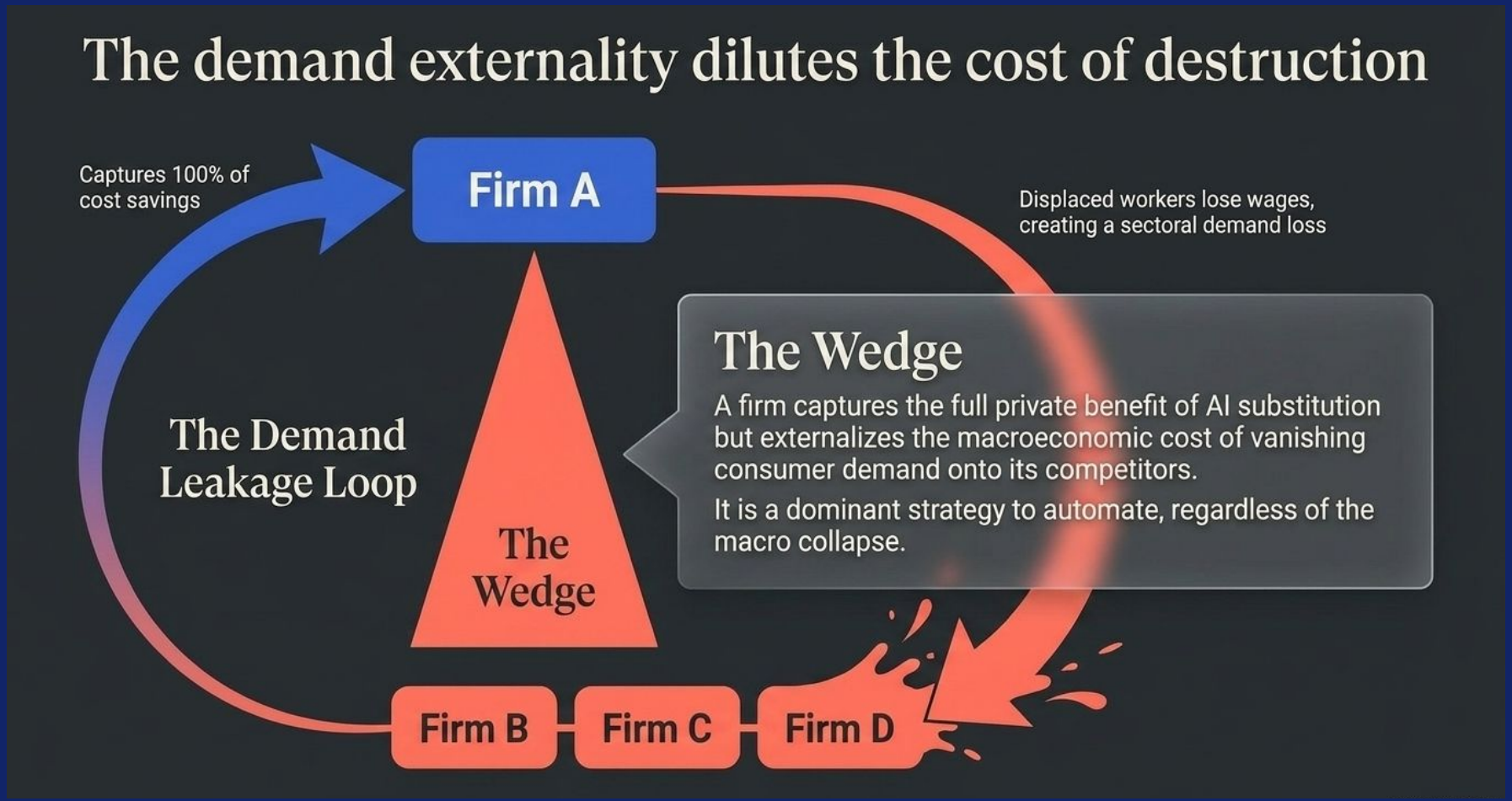
The Standard Intuition

Standard economic intuition assumes forward-looking firms will naturally brake before automation becomes self-destructive. If vanishing paychecks mean vanishing customers, profit-maximizing behavior should enforce a natural speed limit.

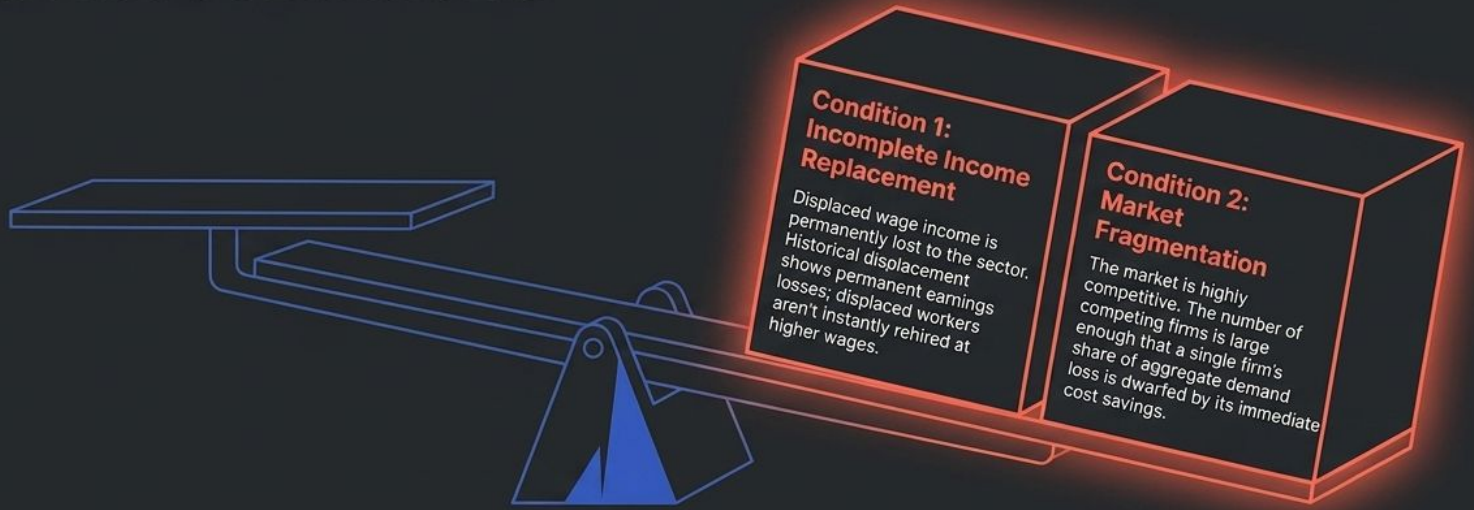
The Reality Check

Public discourse calls this technological determinism. But the math reveals it is actually a structural market failure. Rational firms are actively choosing a path that harms both workers and shareholders.

The demand externality dilutes the cost of destruction

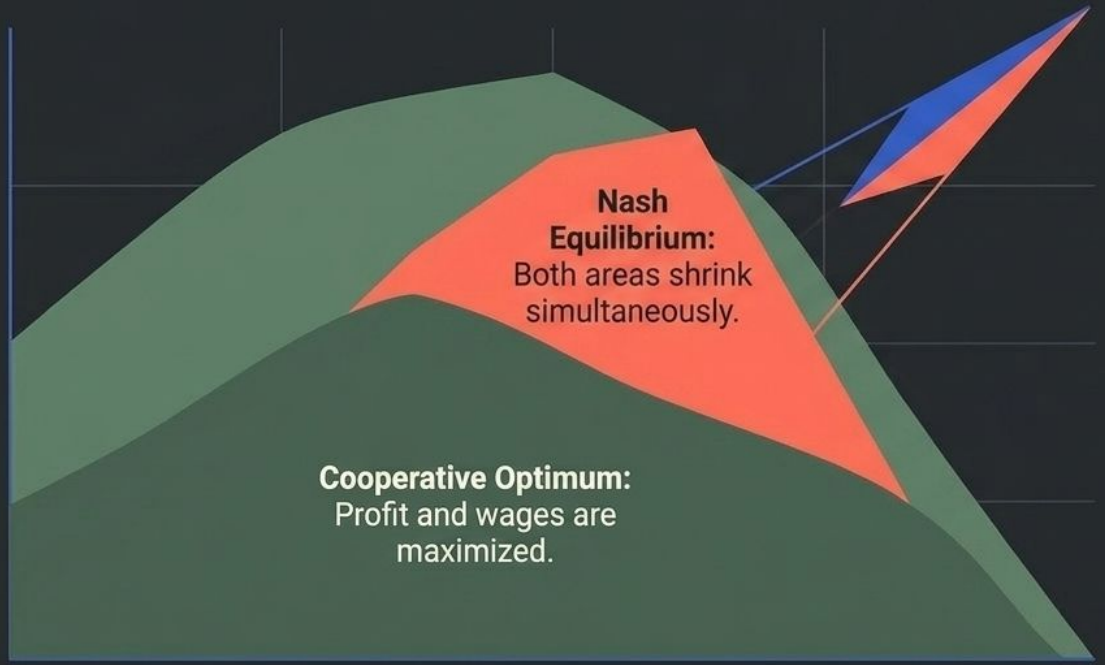


The trap only activates under two specific market conditions



The tech sector currently features rapid displacement of entry-level cognitive roles combined with hyper-fragmented AI integration. Both conditions are heavily met.

Over-automation is a deadweight loss, not a wealth transfer



The Reality

Firms in competitive markets are currently automating at roughly twice the collectively efficient rate.

The Mechanic

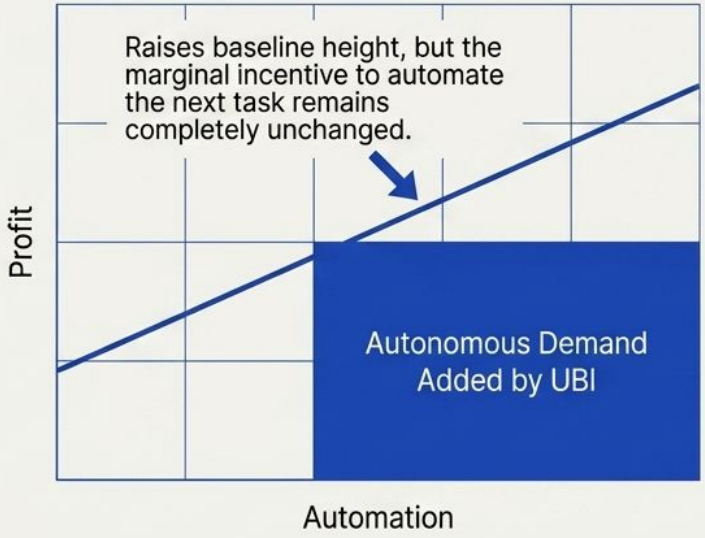
Because competitive pricing dilutes each firm's share of the demand loss, total aggregate profit falls below what firms would earn if they collectively agreed to restrain AI adoption. This is a net destruction of value harming both shareholders and labor.

Evaluating the organizational and policy interventions

Which of these actually operates on the correct margin to stop the race to the cliff?

Safety nets raise the floor, but leave the cliff intact

The False Floor of UBI



Case Study: Amazon Upskilling

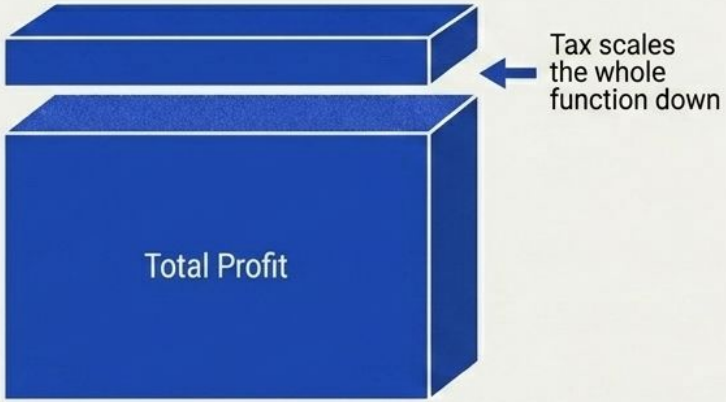
The Action: Amazon's Upskilling 2025 transitioned displaced logistics workers to cloud/AI roles.

The Result: Raised internal income replacement from 0.40 to 0.90 for participants.

The Limit: Shrinks the demand loss per task, but a residual externality remains unless displaced income is 100% replaced and instantly recycled.

The illusion of broad redistribution and worker equity

Capital Taxation



A tax on profit scales the whole function down. Because it applies to the total, it cancels out in the derivative. It changes levels but not the margin.

Case Study: Mondragon Worker Equity

The Action: Spanish cooperative where workers are owners.

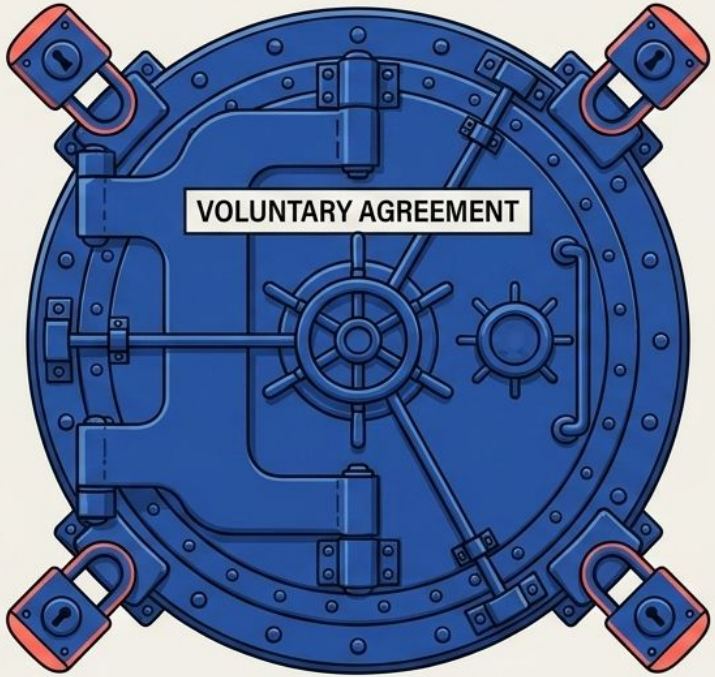
The Reality: When they automated, profits recycled to workers, drastically shrinking the wedge.

The Flaw: Because workers spend less than 100% of their income inside the cooperative, demand still leaks to external competitors. The multilateral externality survives.

Why voluntary restraint inevitably breaks down

Dominant Strategy

Immediate cost savings make full automation the strictly dominant move regardless of rivals.



Multilateral Diffusion

Damage is spread too thinly across firms to motivate bilateral negotiations, but is fatal in aggregate.

Non-contractibility

Automation rates are internal black-box decisions. Rivals cannot verify or enforce compliance.

Irreversibility

AI integration requires massive sunk costs. You cannot realistically undo a deviation once deployed.

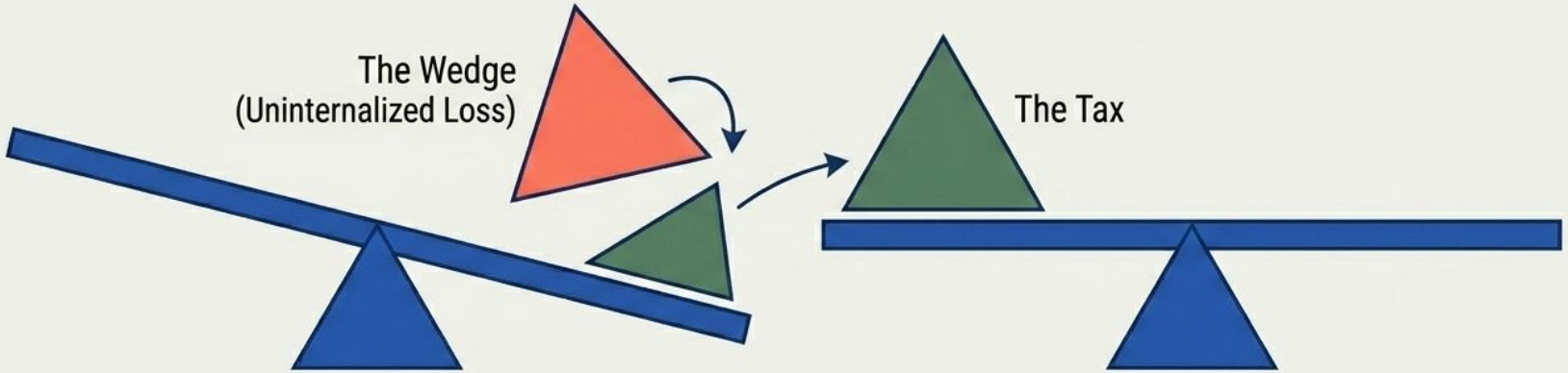
The market failure is incentive-incompatible with private treaties.

Diagnosing the interventions

Policy Tool	Raises Income Floor?	Addresses Cross-Firm Externality?	Alters Marginal Incentive?	Self-Enforcing?
Upskilling	●	◐	○	●
UBI	◐	○	○	●
Capital Tax	○	○	○	●
Worker Equity	●	◐	○	○
Voluntary Agreement	○	◐	○	○
Pigouvian Automation Tax	●	●	●	●

Only a targeted automation tax operates on the exact margin needed to align private incentives with collective welfare.

The Pigouvian Solution: Taxing the uninternalized demand loss



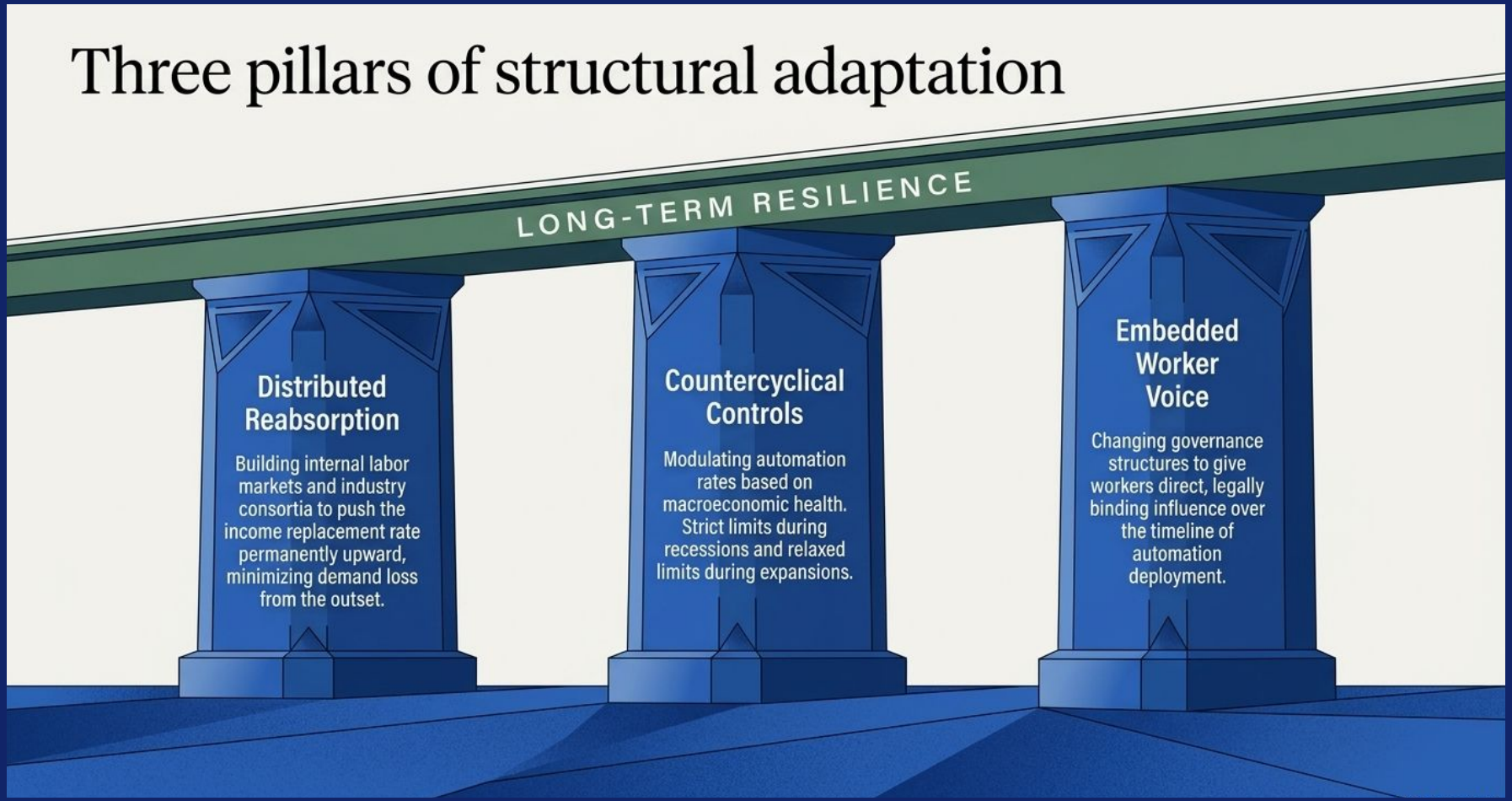
The Mechanic

A per-task tax is levied on automated tasks. Setting the tax equal to the demand loss forces the firm to physically pay for the damage it imposes on rivals. Private cost now equals social cost.

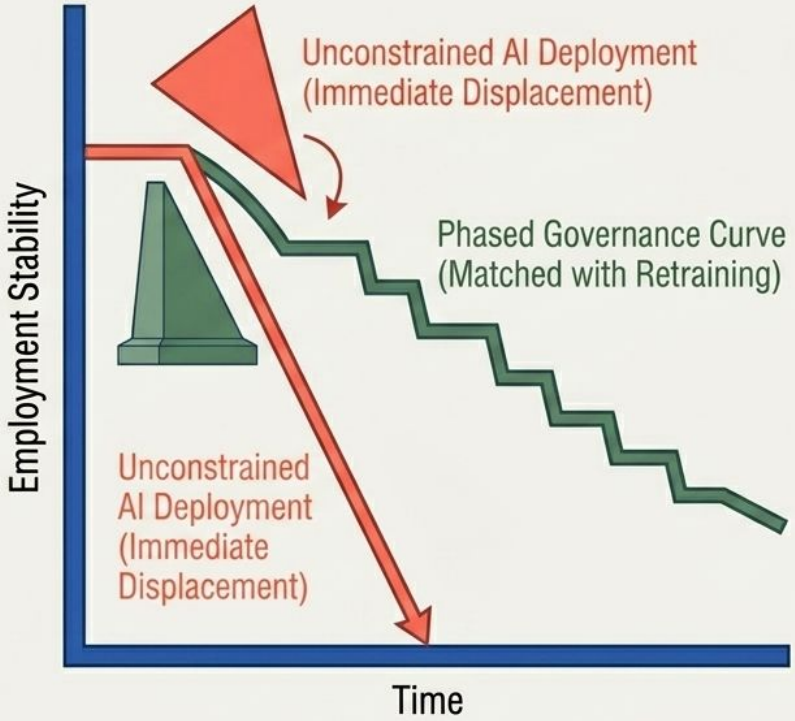
The Magic Loop (Denmark Flexicurity Model)

The tax funds rapid state-sponsored retraining. Retraining raises income replacement. Higher income replacement lowers aggregate demand loss. Lower demand loss means the required tax rate naturally shrinks over time. It is a transitional correction, not a permanent penalty.

Three pillars of structural adaptation



Internalizing the shock through governance



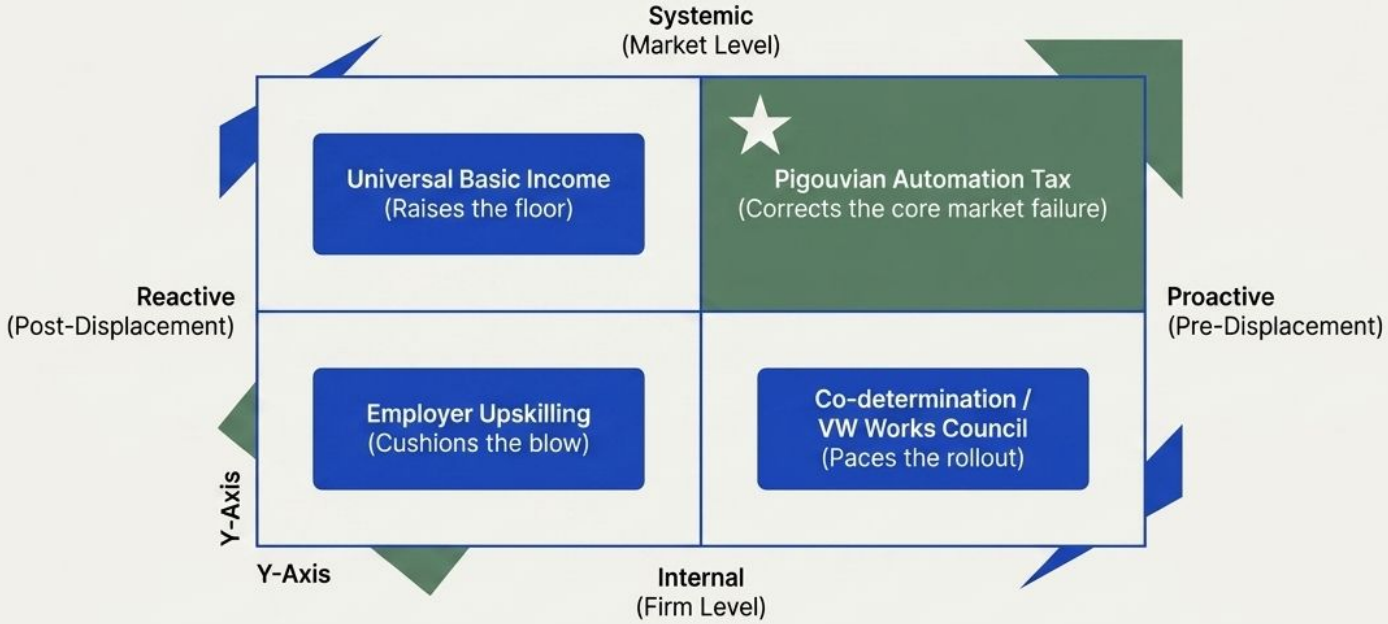
Case Study: Volkswagen's Works Council

The Intervention: VW introduced AI-driven assembly optimization. Rather than immediate displacement, the Works Council utilized legal co-determination rights to negotiate a binding 5-year phase-in.

The Result: The delay allowed displaced assembly workers to access apprenticeships in battery-system integration (a high-demand EV role).

The Impact: The phased rollout kept income replacement above 0.85, eliminating the demand externality while preserving the core productivity gains of the AI.

The Synthesis: A layered portfolio approach



Rationality and perfect foresight are not enough to stop the race. Only proactive institutional design—pairing corrective taxation with governance and capability building—can align private incentives with collective welfare.