# Data as Inventory

A Conceptual Framework for AI Taxation

A Provocation for Policymakers and Economists

William Peterson | December 2025

## THE CORE ARGUMENT

Current AI tax proposals focus on taxing activity (compute, transactions, robot hours). But the real source of durable competitive advantage is the stock—accumulated data that compounds over time.

*The insight:* Compute is like a factory running—anyone can rent it. Data is like inventory in a warehouse—it accumulates and creates lasting advantage. We tax physical inventory. Why not data inventory?

*The purpose:* This paper does not offer a legislative proposal. It asks economists and tax scholars to investigate whether treating retained AI-relevant data as taxable inventory could better align private incentives with public goals.

## Why Now: The Policy Moment

AI policy is at an inflection point. The federal government is actively shaping the regulatory landscape:

- The December 2025 Executive Order on AI establishes a "minimally burdensome national policy framework" while pushing back on state-level regulation
- Over 1,000 AI-related bills were introduced across U.S. states in 2025, creating a patchwork of approaches
- The OECD continues work on digital taxation through the Pillar One and Pillar Two frameworks, grappling with how to tax value created through data and digital presence
- Antitrust scrutiny of major technology platforms has intensified, with market concentration a bipartisan concern

Yet the taxation debate remains stuck on familiar categories—robot taxes, compute levies, profit windfalls—none of which address the structural source of AI advantage: accumulated data. This paper offers a new lens at a moment when policymakers are actively seeking frameworks.

## The Problem: Taxing Activity When the Asset Is Stock

Most AI taxation proposals target flows:

Robot taxes aim to address automation-driven labor substitution, but face severe attribution challenges and may discourage productivity-enhancing innovation.

Compute taxes target GPU hours or energy consumption, but compute is fungible—it can be rented or relocated—and such levies do little to address durable concentration.

Profit and windfall taxes attempt to capture economic rents after the fact, but they require contentious valuation and remain vulnerable to accounting strategies.

None of these approaches targets the accumulation of data as a structural source of advantage. The moat isn't the model running—it's what the model learned and retained.

## The Insight: Data Behaves Like Capital

Unlike physical inventory, which depreciates, data often appreciates—strengthening the case for treating it as taxable capital rather than ignoring it entirely. It differs fundamentally from labor and compute: once collected, it can be reused across model generations, repurposed for new tasks, and combined with additional signals to increase marginal value.

Data is durable: unlike compute cycles consumed in processing, data persists. It is reusable: the same dataset can train multiple models across generations. It is accumulative: early data advantages compound over time. And it is path-dependent: data collected early shapes optimization targets and system behavior, reinforcing incumbency.

These characteristics cause retained data to behave economically like inventory or intangible capital. A growing body of research documents that intangible-intensive firms earn persistently higher returns, and that market concentration has increased where data and IT investments are highest.

The analogy is simple: We tax physical inventory sitting in a warehouse. Why not data inventory sitting on servers?

## What a Data-Inventory Approach Might Include

This paper does not prescribe specific rates or mechanisms—those require expertise beyond this conceptual exercise. However, a data-inventory framework might include:

- *Retention escalators*: Longer retention creates higher liability, discouraging indefinite hoarding.

- *Credits for deletion or release*: Incentives for data minimization, anonymization, or public contribution
- *Volume-based assessment*: Liability tied to quantity of retained AI-relevant data (e.g., terabyte-years)
- *Functional tiering*: Different treatment for training data, behavioral logs, embeddings, etc.
- *Exemptions*: Protections for nonprofit research, small entities, accessibility technologies

The goal would not be revenue maximization but incentive alignment—making the social costs of data hoarding visible in private decision-making, while preserving innovation incentives.

## Open Questions Requiring Further Expertise

This conceptual framework raises questions that require deeper investigation:

- *Measurement*: How would firms report data inventories? How would auditing work?
- *Definition*: What distinguishes "AI-relevant data" from ordinary business data? For purposes of this framework, AI-relevant data might include structured datasets used for model training, behavioral logs, embeddings, and synthetic outputs retained for future use.
- *Jurisdiction*: How would cross-border data storage be handled? Can destination-based principles apply?
- *Evasion*: Could firms embed data into model weights to avoid inventory accounting?
- *Innovation effects*: Would this slow beneficial AI development, or level the playing field?
- *Empirical grounding*: Can economists isolate data's specific contribution to excess returns?
- *International coordination*: How would this interact with OECD digital taxation frameworks?

These are not objections—they are invitations for the economists, tax scholars, and technologists who have the expertise to develop answers.

## Scope and Non-Goals

This paper does not propose regulating AI models, limiting compute, capping profits, or restricting deployment. It does not address national security uses of AI or advocate changes to export controls.

The framework is intended to complement—not replace—existing antitrust, privacy, and competition policy tools. It offers a different lens: viewing data accumulation as an economic phenomenon with structural consequences that current tax frameworks do not address.

## Conclusion

AI taxation debates have focused on visible outputs—compute, transactions, profits. Retained data is a less visible but more durable source of compounding advantage.

This paper's contribution is conceptual reframing: a provocation for policymakers and economists to investigate whether we have been looking at the wrong part of the AI value chain. If the answer is yes, a data-inventory approach—properly designed by those with the relevant expertise—could discourage hoarding, weaken entrenched dominance, improve privacy outcomes, and preserve incentives for innovation.

The question is worth asking. The answer requires further work.

## Selected References

Autor, D., Dorn, D., Katz, L., Patterson, C., & Van Reenen, J. (2020). The Fall of the Labor Share and the Rise of Superstar Firms. Quarterly Journal of Economics, 135(2), 645–709.

Bessen, J. (2017). Information Technology and Industry Concentration. Boston University School of Law, Law and Economics Research Paper No. 17-41.

De Loecker, J., Eeckhout, J., & Unger, G. (2020). The Rise of Market Power and the Macroeconomic Implications. Quarterly Journal of Economics, 135(2), 561–644.

Haskel, J. & Westlake, S. (2017). Capitalism Without Capital: The Rise of the Intangible Economy. Princeton University Press.

Philippon, T. (2019). The Great Reversal: How America Gave Up on Free Markets. Harvard University Press.

## Author's Note

This paper was drafted with the assistance of AI-based language tools. The author is not a tax expert or AI specialist; the contribution here is conceptual rather than technical. All arguments and conclusions are those of the author.

## Contact

William Peterson

Wpeterson.papers@proton.me