The Data–Money Wedge:
A Coupled-Systems Diagnosis of Economic Fragmentation under Artificial Intelligence

(Research Agenda)

ABSTRACT

The macroeconomic debate over artificial intelligence is conducted in the language of aggregates, centering on how much AI will raise total factor productivity and how quickly. We argue that this framing misses the economically consequential phenomenon. The digital and the real economy are neither separate systems to be summed nor a single aggregate; they are one coupled system fragmenting along the axis that joins them, the interface at which data produced by individuals is, or fails to be, converted into money. We characterize this wedge as the joint product of two reinforcing fragmentations: a *data fragmentation*, in which data is a strategic input that cannot achieve singleness across jurisdictional, platform, sectoral, provenance, and ownership boundaries; and a *monetary fragmentation*, in which the resulting divergence in value surfaces as a bifurcated price level. The comparative implication is sharp: economies with different data-governance regimes, especially China, the United States, and Europe, should exhibit different wedge dynamics as AI diffusion raises the value of verified human-origin data and shifts settlement toward machine-legible channels. We give the diagnosis a formal skeleton, a two-substrate model from which the bifurcation, the heterogeneous incidence of inflation, the uncompensated conversion wedge, and the dimensional inadequacy of single-instrument monetary policy follow as formal results. We then set out the empirical baseline, four transmission channels, and the scenario fan the model implies, and we specify the falsifiable claim and the research program that would test it. This is a research-agenda paper; it specifies what proving the diagnosis would require rather than offering a new identification strategy.

JEL classification: E31, E58, O33, D83, E25, G20

Keywords: data economy; inflation heterogeneity; monetary fragmentation; singleness of money; artificial intelligence; heterogeneous agents

Rafael Andrezo
Director of Economics
DrumWave
2465 Latham St
Mountain View, CA

André Vellozo
CEO
DrumWave

1. Introduction

The macroeconomics of artificial intelligence has so far been a debate about the size of a single number. Optimistic forecasts project that generative AI will add several percentage points to output growth over a decade. The most careful skeptical estimate concludes that the gains to total factor productivity are real but modest, no more than roughly two-thirds of a percent over ten years (Acemoglu, 2025). The two camps disagree about magnitude, but they share a framing. Both treat the economic significance of AI as a shift in an aggregate, whether productivity, output, or the price level, that can be read off the instruments economies already use to measure themselves.

In this paper we argue that the framing, rather than the magnitude, is the problem. The consequential effect of AI is not a movement in an aggregate but a fragmentation of the system that the aggregate is meant to summarize. The digital economy and the real economy are not two sectors to be added, and they are not a single homogeneous quantity. They are one coupled system, and that system is fragmenting along the seam that joins them: the interface at which data, produced continuously by individuals, is converted into money, or fails to be.

The asymmetry at that seam is our starting point. Machines do not need money; they need data. People produce data and need money. The data that the AI economy runs on is generated, in the ordinary course of economic and social life, by individuals who are not compensated for it in the medium they require. The value created at the interface accrues to whoever controls the aggregation and conversion of data, not to those who produce it. This is not a claim about two civilizations. It is a claim about where, within a single economic system, value is created and where it is realized.

We call the resulting seam the *data–money wedge*, and we characterize it as the joint product of two fragmentations that reinforce one another. The first is a fragmentation of data as an economic input. Data cannot be accepted at par in the way money can, because its value depends on context, provenance, licensing, and use, and it is divided across incompatible jurisdictional, platform, sectoral, provenance, and ownership boundaries. The second is a fragmentation of money in the precise sense of the literature on the singleness of money (Gorton & Zhang, 2023; Bank for International Settlements [BIS], 2025): a nominally unified system functioning as several incompatible sub-systems, with arbitrage blocked. This monetary fragmentation surfaces concretely as a bifurcated price level, in which the same nominal unit purchases divergent real baskets depending on which side of the provenance divide a household stands.

The paper makes four contributions. First, it locates and characterizes the wedge, establishing the substrate asymmetry, the non-singleness of data, and the five axes along which data fragments (Section 3). Second, it gives the diagnosis a formal skeleton: a two-substrate model from which the bifurcated price level, the heterogeneous incidence of inflation, the uncompensated conversion wedge, and the dimensional inadequacy of single-instrument monetary policy follow as formal results rather than assertions (Section 4). Third, it sets out the empirical baseline as it actually stands, the four transmission channels through which the fragmentation propagates, and the scenario fan the model implies (Sections 5 through 7). Fourth, it states the falsifiable claim and the research program that would test it (Sections 8 and 9). The contribution is the diagnosis and the program it implies, not a new identification strategy.

The 2026 NBER AI conference frames the economics of AI in comparative terms, with emphasis on China and the rest of the world. Our diagnosis naturally yields such a comparison. China is a high-state-capacity, high-data-governance, high-localization case; the United States is more platform-led; Europe is more rights- and compliance-led. If the data–money wedge is real, these regimes should not merely differ in AI adoption rates. They should differ in the path by which data value is enclosed, priced, settled, and passed through to households.

The paper proceeds as follows. Section 2 locates the contribution against four literatures. Section 3 develops the diagnosis and is the analytical core. Section 4 presents the two-substrate model. Section 5 establishes the empirical baseline, Section 6 the four transmission channels, and Section 7 the scenario analysis. Section 8 draws out the implications for monetary policy, Section 9 sets the research agenda, and Section 10 concludes. An appendix sketches the heterogeneous-agent extension that the agenda calls for.

We are explicit about what the paper is and is not. It is a research-agenda paper in the sense of Korinek and Suh (2024) and of the agenda set out by Agrawal, Gans, and Goldfarb (2019). It organizes existing results around a diagnosis, gives that diagnosis a formal structure, and identifies what would be required to prove or refute it. It is not an empirical paper; we do not have proprietary data or a causal identification strategy, and we do not claim one.

2. Related literature

The diagnosis sits at the intersection of four literatures. Each supplies an indispensable component, and none assembles the whole. We take them in turn, stating in each case what

the literature establishes and where our claim departs from it, and then describe the join that constitutes our contribution.

2.1 Data as a non-rival, externality-laden input

The first literature establishes what kind of economic object data is. Jones and Tonetti (2020) show that data is non-rival: unlike a machine or a barrel of oil, the same dataset can be used by many producers at once without being depleted, which creates increasing returns and makes the allocation of property rights, rather than scarcity, the central economic question. Their central result is that who owns data determines whether it is used efficiently. When firms own data they tend to hoard it for fear of eroding competitive advantage or breaching privacy, so consumer ownership can be welfare-superior. Farboodi and Veldkamp (2021) embed data in a growth model in which data is a byproduct of economic transactions and feeds back into productivity, generating a data-driven accumulation process distinct from ordinary capital, and Veldkamp and Chung (2024) survey how this byproduct property reshapes aggregate dynamics.

A second strand within this literature concerns the peculiar pricing of data. Bergemann, Bonatti, and Gan (2022) show that because one individual's data is statistically informative about many others, a platform can infer an individual's characteristics from other people's data, which collapses the price any single individual can command and constitutes a data externality. Acquisti, Taylor, and Wagman (2016) survey the economics of privacy and document how context-dependent and intermediated the valuation of personal data is. Arrieta-Ibarra, Goff, Jiménez-Hernández, Lanier, and Weyl (2018) and Posner and Weyl (2018) reframe data as labor rather than capital: users are the producers of a valuable input, the prevailing arrangement treats their contribution as free, and that framing suppresses the

market that would otherwise compensate them and routes the surplus to whoever controls the data.

What this literature does not do is connect the data input to the monetary system. It characterizes data as a factor with externalities and contested ownership, but it stops at the level of the firm and the data market. It does not model the interface at which the value of data is, or is not, converted into money for its producers, and it does not draw the consequence for the price level. The non-singleness of data, implicit in the heterogeneity and context-dependence this literature documents, is precisely the bridge we use to carry these insights to the monetary side.

2.2 Automation and the macroeconomics of AI

The second literature concerns what AI does to production and distribution. The task-based framework of Acemoglu and Restrepo (2018) decomposes the effect of automation into a displacement effect, which removes tasks from labor, and a reinstatement effect, which creates new ones, with the net effect on wages and the labor share depending on their balance. In subsequent work the same authors attribute between half and two-thirds of the change in the U.S. wage structure between 1980 and 2016 to automation that displaced workers specialized in routine tasks (Acemoglu & Restrepo, 2022). Moll, Rachel, and Restrepo (2022) extend the logic to wealth, showing that because automation raises the return to the capital that does the automating, it widens wealth as well as income inequality.

The same framework disciplines forecasts of AI's aggregate effect. Acemoglu (2025) argues that only a fraction of tasks are cost-effectively automatable within a decade, so the implied gain to total factor productivity is modest. Korinek and Suh (2024) trace scenarios for the

transition to more general AI, in which wages can rise and then fall sharply if machine capability comes to span the full range of human tasks, and Brynjolfsson (2022) argues that the private incentive to build human-substituting rather than human-augmenting systems concentrates economic power even where augmentation would raise output more.

This literature treats money as a neutral veil. It analyzes quantities, factor shares, and real wages, and it does not contemplate a fragmentation of the monetary substrate or a bifurcation of the price level. Its distributional results, moreover, are organized around the labor-versus-capital margin. Our wedge sits upstream of that margin, at the point where the data input fails to convert into money for the households that produce it, and our model is consistent with the Moll, Rachel, and Restrepo capital-share result as a limiting case in which the conversion wedge is large and broadly uniform (Implication 3).

2.3 The heterogeneity of inflation incidence

The third literature documents that inflation is not experienced equally. Using retail scanner data, Kaplan and Schulhofer-Wohl (2017) show that households face systematically different inflation rates, with dispersion that is large and persistent rather than transitory noise. Jaravel (2019) demonstrates that product innovation is concentrated in the segments serving higher-income households, so that the variety-adjusted inflation they face is lower, an inflation inequality that favors the rich; his subsequent review synthesizes the measurement and the causes (Jaravel, 2021). Argente and Lee (2021) find that during the Great Recession lower-income households experienced materially higher cost-of-living growth than richer ones.

This literature establishes the fact of heterogeneous incidence and traces it largely to income-correlated differences in consumption baskets and in exposure to product variety. It treats that heterogeneity as a distributional property of an otherwise unified price level. We supply a new and structural source of basket dispersion specific to the AI economy: the provenance axis of data fragmentation drives a wedge between the price path of authenticated, human-origin goods and that of synthesizable goods, and households differ in how their baskets straddle that divide. In our model this is Proposition 2: the cross-sectional variance of household inflation scales with the squared gap between authenticated- and synthesizable-good inflation, conditional on heterogeneity in household expenditure shares.

2.4 Singleness of money and heterogeneous transmission

The fourth literature supplies the monetary concept on which the diagnosis turns. Gorton and Zhang (2023) articulate the singleness of money: a monetary system functions only if each unit is accepted at par against every other, with no questions asked, and they show that private monies which fail this test, from the free-banking era to modern stablecoins, fragment into instruments trading at varying and uncertain values. The Bank for International Settlements (2025) frames singleness, elasticity, and integrity as the three properties a sound monetary system must have, and identifies tokenization as a development that puts singleness under pressure. This is the property we transpose to the data side: data is precisely the input that cannot be accepted at par across contexts.

A parallel strand shows that even within a unified monetary system, transmission is distributionally uneven. Kaplan, Moll, and Violante (2018) show that in a heterogeneous-agent New Keynesian economy, monetary policy works mainly through indirect general-equilibrium channels and depends on the distribution of marginal propensities to consume,

a mechanism representative-agent models cannot see. Auclert (2019) identifies the redistribution channel through which interest-rate changes transfer resources across households with different balance-sheet exposures. The empirical sign of these effects is genuinely contested: Lenza and Slacalek (2024) find that euro-area quantitative easing compressed income inequality through employment while leaving wealth inequality little changed, whereas Andersen, Johannesen, Jørgensen, and Peydró (2023), using Danish administrative data, find that monetary easing raised income inequality through non-labor channels.

What this literature assumes away is our object. It takes the monetary substrate as unified and asks how policy propagates through it. We ask what happens to the substrate itself when the input feeding the economy fails singleness, and we use the heterogeneous-agent machinery to explain why a single-index, mean-targeting instrument may fail to stabilize dispersion when the bifurcation gap is left unchanged, which is the content of Proposition 5.

2.5 The contribution: the join

Our contribution is to connect these four literatures rather than to add a result inside any one of them. The data literature supplies the non-singular input and the uncompensated producer; the singleness literature supplies the monetary property that data violates; the inflation-heterogeneity literature supplies the observable symptom; and the heterogeneous-agent literature supplies the reason aggregate instruments miss it. The diagnosis is what emerges when they are joined: a single coupled system fragmenting at the interface between data and money, with consequences that are distributional and that the standard measurement apparatus is built to overlook. This places the paper within the research

agenda for the economics of artificial intelligence set out by Agrawal, Gans, and Goldfarb (2019), as a diagnosis and a program rather than an estimate.

The position our framework most directly confronts is Acemoglu (2025), and we do not contest it. If the aggregate productivity gain from AI is modest, our claim is untouched, because it is orthogonal. A fragmentation operates on the distribution of prices and on the adequacy of measurement, not on the level of the aggregate. A small aggregate gain is fully consistent with a large and consequential fragmentation of the system that produces it, and a fragmentation is most dangerous precisely when the aggregate looks unremarkable and the instruments register nothing amiss.

3. The wedge: one system fragmenting at the data–money interface

This section sets out the diagnosis. It is the analytical core of the paper, and everything that follows is downstream of it. We establish the substrate asymmetry (Section 3.1) and the non-singleness of data (Section 3.2), develop the five axes along which data fragments (Section 3.3), locate the wedge among those axes (Section 3.4), characterize the reinforcing feedback that binds the two sides (Section 3.5), state the standard of proof appropriate to a research-agenda paper (Section 3.6), and explain why aggregate measurement misses the phenomenon (Section 3.7). The formal counterpart of the argument is the model of Section 4.

3.1 The substrate asymmetry

Begin with what each side of the system requires. Human participants in the economy require money, the medium in which they are paid, hold savings, and settle obligations. The AI production process requires data, on which models are trained and operated and whose

quality and provenance bound what those models can do. The asymmetry is that the data the second process requires is produced, as an ordinary byproduct of economic and social life, by the same individuals who require money, and those individuals are not, under present arrangements, compensated in money for producing it.

Two established results make this precise. First, data is non-rival (Jones & Tonetti, 2020): the same data can be used by many processes at once, so its social value can be large even where its private price is near zero. Second, the value individuals create by producing data is more naturally understood as labor than as a free gift of capital (Arrieta-Ibarra et al., 2018). Treating it as free suppresses the market that would otherwise compensate it and routes the surplus to whoever controls the data. The asymmetry is therefore not that two economies run on two currencies. It is that within one economy the input that increasingly drives production is generated by individuals who are paid neither in that input nor, for producing it, in the money they need. In the model of Section 4 this is the gap between the value of the data households generate and the fraction of that value, governed by the interface-access parameter, that returns to them as income.

3.2 Why data cannot behave like money: non-singleness

One might try to resolve the asymmetry by treating data as a kind of money, a new monetary base for the AI economy. That move fails, and understanding why it fails reveals the source of the wedge. Money must satisfy the singleness property: it is accepted at par, with no questions asked, whatever form it takes (Gorton & Zhang, 2023; BIS, 2025). A dollar is a dollar, and one does not perform due diligence on it before accepting it. Data is the opposite. Its value is entirely a matter of questions asked, since a dataset is worth radically different

amounts depending on its provenance, the license under which it is held, the jurisdiction that governs it, the platform that captured it, and the use to which it will be put.

This non-singleness is not a friction to be engineered away. It is constitutive of what data is. Because data cannot be accepted at par across contexts, it cannot integrate into a single market the way money does, and so it fragments. The deep reason the wedge exists is that the input on the data side fails the very property that lets the money side function as a unified system. The two sides cannot clear against each other at par, and value accumulates on one side of the interface without converting to the other. Section 4 formalizes this as a fragmentation wedge that reduces the effective data input below the raw total that would be available if data were singular, with money corresponding to the limiting case of zero fragmentation.

3.3 The five axes of data fragmentation

Data fragments along at least five axes. Each blocks the arbitrage that would otherwise equalize the value of data across contexts, and each is observable and widening. We treat them in turn, since they are the empirical substance of the non-singleness claim and the raw material for the feedback of Section 3.5.

3.3.1 Jurisdictional fragmentation

Data is governed by incompatible national and regional regimes, so the same data carries different rights, obligations, and usable value depending on where it is held and processed. The European Union's General Data Protection Regulation, California's privacy statutes, China's Personal Information Protection Law, and India's data-protection framework impose mutually inconsistent requirements, and an expanding set of data-localization mandates requires that data be stored or processed within national borders. This is a literal

barrier to par-conversion: a unit of data lawfully usable in one jurisdiction may be unusable, or unlawful to move, in another.

The barrier is large and growing. The Information Technology and Innovation Foundation documents that the number of data-localization measures worldwide more than doubled between 2017 and 2021, from 67 to 144, spread across 62 countries, with China, India, Russia, and Turkey the most restrictive (Cory & Dascoli, 2021). Such measures have been found to raise data-management costs by between 15 and 55 percent and to push up prices for the downstream industries that depend on data. Jurisdictional fragmentation thus not only blocks conversion; it has a measurable price effect, the first hint of the link between data fragmentation and the price level that Section 4 makes precise.

3.3.2 Platform fragmentation

Data captured inside one platform's systems is inaccessible to its competitors, and the same user, generating data continuously across many platforms, holds transferable property rights in none of it. The same nominal resource therefore has different value depending on which silo it occupies, and there is no market on which a unit of data in one silo trades at par against a unit in another. The closing of previously open interfaces has made the walls explicit. Reddit began charging for commercial interface access in 2023 and, according to its 2024 registration statement on Form S-1, had entered data-licensing arrangements with an aggregate contract value of 203 million dollars (Reddit, Inc., 2024); contemporaneous reporting attributed roughly 60 million dollars per year to an agreement with one major model developer and about 70 million to another.

These transactions reveal the fragmentation directly. Data that is nominally the same, namely user-generated text, commands sharply different prices depending on which

platform controls it and is negotiated bilaterally, deal by deal, rather than clearing at a single price. A reported ecosystem of comparable agreements, from a roughly 250-million-dollar arrangement between a major news publisher and a model developer to one-time payments for academic archives, confirms that the price of data is platform- and corpus-specific. The very existence of a bespoke licensing market is evidence of non-singleness: singular goods do not require deal-by-deal price discovery.

3.3.3 Sectoral fragmentation

Data is further partitioned by sector. Medical, financial, employment, and biometric data trade in separate regulated markets, each with its own access rules, permissible uses, and pricing, governed by distinct statutes. Health data, credit data, and biometric identifiers are subject to regimes that do not map onto one another, and combining them into a single training corpus is legally and operationally costly. This is the data analogue of the conversion costs that sustain price differences in segmented asset markets: arbitrage is not impossible, but it is expensive enough that values do not equalize across sectors.

The consequence is that the effective value of a given quantity of data depends on the sector it sits in and on whether it can be lawfully combined with data from other sectors. A model that could integrate clinical, financial, and behavioral data would be more capable than the sum of models confined to each silo, but the regulatory partition prevents that integration from being priced and traded in a unified market. Sectoral fragmentation is, in this sense, a standing destruction of the complementarities that non-rivalry would otherwise deliver.

3.3.4 Provenance fragmentation

Provenance is the axis on which fragmentation is in kind rather than in location, and it is the one that maps most directly onto the price level. Verified human-generated data commands

a premium that has risen as model-generated content has proliferated, while synthesizable data deflates in value as its supply expands toward the limitless. The economic basis for the premium is established in the technical literature: Shumailov et al. (2024) show that models trained indiscriminately on recursively generated content suffer model collapse, an irreversible degradation in which the tails of the original distribution disappear. Verified human data is the input that prevents this collapse, and its scarcity value rises precisely as synthetic content floods the available corpus.

Human-origin and machine-origin data are therefore no longer fungible, and the market has begun to price the distinction explicitly through provenance and authentication standards and through licensing arrangements that specifically value human-authored corpora. This is fragmentation in the strong sense: two units of data that are superficially identical command different prices because of where they came from. In the model of Section 4 this axis is the engine of the bifurcated price level, since the rising premium on authenticated provenance is what drives the relative price of authenticated goods above that of synthesizable ones.

3.3.5 Ownership fragmentation

The fifth axis is the one along which the conversion to money actually fails, and it is where, in Section 3.4, we locate the wedge. The prevailing arrangement captures data under terms of service that transfer operational control without compensation, splitting the legal owner of the data, nominally the individual, from its operational controller, the platform. Following Arrieta-Ibarra et al. (2018) and Posner and Weyl (2018), the productivity surplus generated by individuals' data accrues to a small number of corporate balance sheets rather than to the individuals who are its proximate source.

A single recent transaction illustrates the wedge cleanly. When a large academic publisher licensed its archive of previously published research to a model developer for a reported 23 million dollars, the payment flowed to the publisher, and questions were raised as to whether the individual researchers whose work constituted the corpus received any compensation at all. The pattern generalizes: the value created at the data-money interface is realized, but it is realized for the controller rather than the producer. The concentration of data-derived value in the balance sheets of a few firms, against the absence of any corresponding data income in household accounts, is the macroeconomic signature of ownership fragmentation, and it is the axis along which the wedge is widest because here the conversion does not merely misprice, it fails outright.

These axes are not independent, and their interactions amplify rather than dampen the fragmentation. Jurisdictional barriers interact with platform silos when platforms are nationally regulated; sectoral partitions interact with provenance when authentication is sector-specific; and ownership fragmentation compounds all the others, since the absence of individual property rights in data is what allows platform, sectoral, and jurisdictional barriers to be drawn around it without the producers' consent. The result is that the fragmentation does not self-correct toward a single price for data. It is, on the contrary, self-reinforcing, which is the subject of Section 3.5.

3.4 Locating the wedge

Among the five axes, the wedge between value created and value realized sits most sharply at the ownership axis, with the provenance axis as its bridge to the price level. The point of contact between the data side and the money side of the system is the moment at which data produced by an individual would convert into money for that individual and, under present

arrangements, does not. The value is real and is realized; it simply accrues to the operational controller rather than to the producer. The ownership axis is therefore where the wedge is widest, because it is the axis along which the conversion fails outright rather than merely being mispriced. Two adjacent readings, one locating the wedge at the aggregate non-singleness of data and one at the provenance premium alone, imply the same downstream consequence, a value divergence the monetary system cannot arbitrage away, so the argument that follows is robust to either.

3.5 The reinforcing feedback

The two fragmentations are not parallel trends. They reinforce one another, and that mutual reinforcement is what makes the two a single coupled system rather than two correlated phenomena. Provenance fragmentation, the rising premium on verified-human data and the deflation of synthesizable data, pushes the relative price of authenticated goods up and that of digitally substitutable goods down. That divergence is the bifurcated price level, a monetary fragmentation in which the same nominal unit no longer purchases a common real basket. The bifurcated price level in turn raises the return to controlling and enclosing verified-human data, which deepens the provenance and ownership fragmentation on the data side. The data fragmentation feeds the monetary fragmentation, and the monetary fragmentation feeds back into the data fragmentation.

Section 4 writes this loop as a dynamic system and shows that it is self-amplifying when the product of the cross-sensitivities exceeds the rate at which competition and regulation erode enclosure, and self-correcting otherwise. Which side leads, whether the data fragmentation is the driving margin or the monetary fragmentation is, is a question of dynamics that

requires additional structure on shocks, timing, and information flow to resolve, and we return to it as the first item of the research agenda in Section 9.

3.6 What would prove the diagnosis

Because this is a research-agenda paper, proving the fragmentation does not mean estimating it with a new identification strategy. It means specifying a falsifiable claim and the observations that would bear on it. The claim is that conventional aggregate measures, headline and core CPI, the standard monetary aggregates, and measured total factor productivity, systematically miss heterogeneous effects in the specific patterns the fragmentation predicts: a widening gap between the price paths of authenticated and synthesizable baskets, a rising share of economic value settled or stored outside the instruments the aggregates capture, and a divergence between measured productivity and the distribution of its returns.

The diagnosis is refutable. If, through a period of rapid expansion in AI production and in data enclosure, the aggregates continue to behave as standard unified-substrate models predict, with the authenticated and synthesizable baskets failing to diverge, value not migrating outside the measured system, and the distribution of returns tracking measured productivity, then the hypothesis fails. If instead the aggregates diverge from household-level and channel-level reality in the patterns the model of Section 4 predicts, the hypothesis is supported. Specifying that test, and constructing the data that would run it, is the burden of the research agenda, and the principal task we leave to future work.

3.7 Why aggregate analysis misses it

A final point bridges to the policy discussion. The reason the fragmentation is not already a settled empirical fact is that the measurement apparatus was built for a unified substrate. A

single price index presumes a common basket; a single monetary aggregate presumes that value settled outside it is negligible; a single productivity residual presumes that the gains it measures are distributed through the same system that produced them. Each presumption is exactly what fragmentation violates. The effects are therefore not absent from the data; they are dispersed below the level at which the standard instruments aggregate, for the same reason that heterogeneous-agent models detect transmission that representative-agent models cannot (Kaplan, Moll, & Violante, 2018; Auclert, 2019). The fragmentation is structured so as to be invisible to the tools currently pointed at it, which is why Section 8 treats the measurement of the bifurcation as a precondition for any policy response.

4. A two-substrate model

This section gives the diagnosis a formal skeleton. The model is deliberately stylized. Its purpose is not to deliver calibrated magnitudes but to show that the wedge, the bifurcated price level, the heterogeneous incidence, and the dimensional inadequacy of conventional policy follow from a small set of assumptions rather than being asserted. We solve the statics in closed form and characterize the dynamics up to a local stability condition, and we are explicit at the end about what the model delivers and what it defers.

4.1 Environment

There is a unit continuum of households indexed by i in [0,1]. The economy produces two final goods. An *authenticated* good A requires verified human provenance and cannot be synthesized; a *synthesizable* good S can be produced with artificial intelligence operating on data. Households consume both goods and differ in two respects that drive every result below: the weight they place on the authenticated good in consumption, and the degree to which the value of the data they generate returns to them as income.

Household i has constant-elasticity-of-substitution preferences over the two goods,

$$U_i = [\ \omega_i^{1/\sigma}\, c_{A,i}^{(\sigma-1)/\sigma} + (1-\omega_i)^{1/\sigma}\, c_{S,i}^{(\sigma-1)/\sigma}\ ]^{\sigma/(\sigma-1)}, \tag{1}$$

where $\sigma > 0$ is the elasticity of substitution between authenticated and synthesizable goods and ω_i in [0,1] is the household-specific weight on the authenticated good. The corresponding price index faced by household i is

$$P_i = [\ \omega_i\, P_A^{1-\sigma} + (1-\omega_i)\, P_S^{1-\sigma}\ ]^{1/(1-\sigma)}, \tag{2}$$

so households with different baskets face different price indices. This is the formal seat of heterogeneous incidence: when P_A and P_S move apart, P_i moves by an amount that depends on ω_i.

4.2 Two substrates: money, data, and non-singleness

The economy has two settlement substrates. Money is the numeraire and satisfies the singleness property: one unit is accepted at par against any other, with no questions asked (Gorton & Zhang, 2023; BIS, 2025). Data is the input that drives synthesizable production, but it does not satisfy singleness. Raw data is generated across cells indexed by the five fragmentation axes of Section 3.3. Let D_{raw} denote the total raw data that would be usable if data were singular, that is, if a unit in any cell were convertible at par into a usable unit in any other. Let θ_j in (0,1] be the fraction of cell-j data value that survives conversion into usable form, and define the effective data input

$$D = (1 - F)\, D_{raw}, \quad \textit{where}\ \ F = 1 - (\ \Sigma_j\, \theta_j\, d_j\ ) \,/\, (\ \Sigma_j\, d_j\ )\ \textit{in}\ [0,1). \tag{3}$$

Equivalently, normalize $D_{raw} = \Sigma_j\, d_j$. Then effective data is $D = \Sigma_j\, \theta_j\, d_j$, and F measures the fraction of raw data value not converted into effective usable data.

The scalar F is the fragmentation wedge: the share of potential data value not converted into effective usable data because data cannot be accepted at par across cells, so a higher F lowers effective data D for a given raw stock. Money corresponds to $F = 0$ by the singleness property. Data has $F > 0$, and this is not a friction to be engineered away but a property of what data is, since value is intrinsically context-, license-, and provenance-dependent.

4.3 Production and the bifurcated price level

Both goods are produced competitively. Labor is mobile across sectors and earns a common wage w. The synthesizable good uses AI capability A, effective data D, and labor, with effective data entering as a non-rival productivity shifter:

$$Y_S = A\, D^{\alpha}\, L_S, \quad \alpha > 0, \tag{4}$$

The same effective data stock can raise productivity across uses without being depleted, the non-rivalry property of Jones and Tonetti (2020). Since labor is the paid variable input and data is appropriated by producers at a price near zero, the competitive unit cost of the synthesizable good is:

$$P_S = w \,/\, (\, A\, D^{\alpha}\,). \tag{5}$$

The authenticated good requires verified human provenance and admits no data substitution. With human productivity B,

$$Y_A = B\, L_A, \quad P_A = w \,/\, B. \tag{6}$$

The relative price of the two goods, which we call the bifurcation, is therefore

$$R \equiv P_A \,/\, P_S = A\, D^{\alpha} \,/\, B. \tag{7}$$

Proposition 1 (Bifurcation).
The bifurcation $R = P_A/P_S = A\,D^{\alpha} / B$ is strictly increasing in AI capability A and in the effective data stock D. Advances in AI and data accumulation lower the price of synthesizable goods relative to authenticated goods, even as the aggregate price index, a weighted average of P_A and P_S, can remain stable or fall.

Proof. Differentiating (7), $\partial R/\partial A = D^{\alpha}/B > 0$ and $\partial R/\partial D = \alpha\, A\, D^{\alpha-1}/B > 0$. Since P_S falls in A and D by (5) while P_A is independent of them by (6), the aggregate index in (2) evaluated at mean weights need not rise. The split widens while the average is quiet. ■

This comparative static concerns AI capability A and effective usable data D. Fragmentation F lowers D for a given stock of raw data and is therefore conceptually distinct from the provenance and enclosure channels that raise the relative scarcity of authenticated goods. The model's static production block shows why AI and effective data accumulation reduce the relative price of synthesizable goods. The fragmentation mechanism operates through the failure to convert raw data into effective data and through the provenance and enclosure feedback characterized below.

Proposition 1 is the model's central claim and the reason an aggregate-first reading of AI misses the phenomenon. The consequential object is the relative price R, not the level of the aggregate. A measured aggregate that looks unremarkable is fully consistent with a large and growing bifurcation underneath it.

4.4 Data production, the conversion wedge, and distribution

Households generate raw data as a byproduct of economic and social activity. The effective data that results, D, has a marginal value product in synthesizable production, which from (4) is the marginal product $\alpha\, Y_S/D$ valued at the price P_S,

$$MVP_D = \alpha\, P_S\, Y_S\, /\, D, \tag{8}$$

positive and rising with the scale of synthesizable production. The question the wedge turns on is who receives this value. Let τ_i in [0,1] be the share of the value of household i's data that returns to that household as money, its access to the conversion interface, and let $\bar{\tau}$ be the mean. Following Arrieta-Ibarra et al. (2018) and Posner and Weyl (2018), the prevailing arrangement sets τ_i close to zero: data is captured under terms of service that transfer control without compensation. The uncompensated transfer from data producers to data controllers, the conversion wedge, is

$$W = (1 - \bar{\tau}) \cdot MVP_D \cdot D. \tag{9}$$

This is the uncompensated transfer out of realized effective data value. A distinct object is the fragmentation loss,

$$L_F = MVP_D \cdot (D_{raw} - D),$$

which captures value not realized because raw data cannot be converted into effective data at par across cells. The paper focuses on W, the distributional transfer from data producers to data controllers, while L_F measures the efficiency loss from non-singleness.

Proposition 2 (Heterogeneous incidence).
Let s_i be household i's expenditure share on the authenticated good and π_A, π_S the inflation rates of the two goods. Household effective inflation is $\pi_i = \pi_S + s_i(\pi_A - \pi_S)$, so its cross-sectional mean and variance are $\bar{\pi} = \pi_S + E[s](\pi_A - \pi_S)$ and $Var(\pi_i) = Var(s)(\pi_A - \pi_S)^2$. The dispersion of lived inflation rises with the dispersion of baskets and, one for one in squared terms, with the bifurcation gap $\pi_A - \pi_S$.

Proof. Linearizing (2), $\pi_i = s_i \pi_A + (1\text{-}s_i)\, \pi_S$. Subtract the mean and take second moments; π_A, π_S are common across households, so only s_i is random, giving $Var(\pi_i) = Var(s)(\pi_A - \pi_S)^2$. The expression shows that dispersion is governed by the squared gap between sectoral inflation rates. Proposition 1 gives the level mechanism behind the relative-price split; whether the inflation gap $\pi_A - \pi_S$ is positive, negative, or widening over a given window is an empirical object. ■

Implication 3 (Wedge and distribution).
Under the accounting assumption that data controllers are capital owners and that uncompensated data value is booked as profits or rents, a low $\bar{\tau}$ shifts realized data value toward capital income. The transfer W in (9) raises the capital share of income and leaves households bearing the bifurcated price index P_i without the offsetting data income that would hedge it. The model is consistent with the automation-driven rise in the capital share (Moll, Rachel, & Restrepo, 2022) as a limiting case in which the conversion wedge is large and broadly uniform, and it adds a distributional margin invisible to factor-share accounting, namely heterogeneity in interface access τ_i.

4.5 The reinforcing feedback and its dynamics

Sections 4.3 and 4.4 are static. The diagnosis turns on a feedback between the two fragmentations (Section 3.5), which requires dynamics. Let the bifurcation evolve with AI capability and with the intensity of data enclosure $e_t = 1 - \tau_t$, the degree to which data value is captured by controllers rather than returned to producers,

$$ln\, R_t = a + b\, A_t + c\, e_t, \quad c > 0, \tag{10}$$

so that enclosing verified-human data raises its scarcity premium and widens the bifurcation. Enclosure, in turn, chases that premium and is restrained by competition and policy,

$$e_{t+1} - e_t = \varphi \cdot ln\, R_t - \delta\, e_t, \quad \varphi > 0, \ \delta > 0, \tag{11}$$

where φ is the sensitivity of enclosure to the return from controlling scarce authenticated data and δ is the rate at which competition and regulation erode enclosure. Substituting (10) into (11) gives a linear law of motion for enclosure, $e_{t+1} = (1 + \varphi\, c - \delta)\, e_t + const$.

Proposition 4 (Feedback stability).
The coupled system (10) and (11) has the local multiplier $\lambda = 1 + \varphi\, c - \delta$. The fixed point is locally stable if $|1 + \varphi c - \delta| < 1$. For small positive adjustment steps, this condition reduces approximately to $\varphi c < \delta$, so the fragmentation is self-correcting when $\varphi c < \delta$ and self-amplifying when $\varphi c > \delta$. The relative magnitudes of c and φ indicate where amplification is strongest. Establishing temporal leadership requires additional assumptions about shocks, timing, or information flow and is therefore left to the research agenda.

Proof (local). Substituting (10) into (11) yields $e_{t+1} = (1 + \varphi c - \delta)\, e_t + (\varphi\, a + \varphi\, b\, A)$, a linear difference equation with multiplier $1 + \varphi c - \delta$. The fixed point is locally stable if and only if $|1 + \varphi c - \delta| < 1$. ■

Remark. This is a local, linearized result under assumed functional forms; the global dynamics and the calibration of (c, φ, δ) are deferred to Section 9.

4.6 Monetary policy: dimensional inadequacy

Suppose the monetary authority has a single instrument and uses it to target an aggregate inflation rate, choosing the economy-wide mean $\bar{\pi}$ to minimize a loss defined over the distribution of household inflation rates relative to a target π^*,

$$L = \int_0^1 (\pi_i - \pi^*)^2 \, di = (\bar{\pi} - \pi^*)^2 + Var(\pi_i). \tag{12}$$

Proposition 5 (Dimensional inadequacy).
A single instrument that controls only the mean $\bar{\pi}$ can set the first term of (12) to zero but cannot affect the second. The minimized loss is the irreducible residual $L^* = Var(\pi_i) = Var(s)(\pi_A - \pi_S)^2 > 0$, which by Proposition 1 grows with the bifurcation. Within this reduced-form setup, a single instrument that moves only the economy-wide mean can stabilize the average but cannot generally stabilize the cross-sectional dispersion. Stabilizing both objects requires either an instrument that affects the bifurcation directly or an additional policy margin.

Proof. Equation (12) is the bias-variance identity $E[(\pi_i - \pi^*)^2] = (E[\pi_i] - \pi^*)^2 + Var(\pi_i)$. The instrument enters only through $E[\pi_i] = \bar{\pi}$; minimizing over it gives $\bar{\pi} = \pi^*$ and residual $Var(\pi_i)$. The result follows under the maintained restriction that the instrument affects $E[\pi_i]$ but not $Var(\pi_i)$ directly. ■

4.7 What the model delivers and what it defers

The model delivers four results in closed form from one set of assumptions: the bifurcation rises with AI capability and data while the aggregate stays quiet (Proposition 1); the dispersion of lived inflation rises with the squared sectoral inflation gap, conditional on basket heterogeneity (Proposition 2); the uncompensated conversion wedge shifts realized data value toward data controllers as a limiting case of the capital-share result (Implication

3); and a single mean-targeting instrument cannot generally stabilize the dispersion of a bifurcating price level (Proposition 5). The feedback that binds the two fragmentations is characterized up to a local stability condition (Proposition 4); the global dynamics and the question of temporal leadership are left to the research agenda.

Three things are deliberately deferred. The global dynamics of the feedback system, beyond the local linearization of Proposition 4, require either stronger functional-form assumptions or numerical solution. The calibration of the structural parameters, in particular the sensitivities (c, φ, δ) and the fragmentation wedge F, requires the provenance-decomposed price series and the conversion-gap measure identified in Section 9. And the general-equilibrium determination of the wage and of the allocation of labor across sectors, held in the background here, would close the model fully. These are the natural contents of the follow-up work, and none is required for the four static results, which stand on the assumptions stated here.

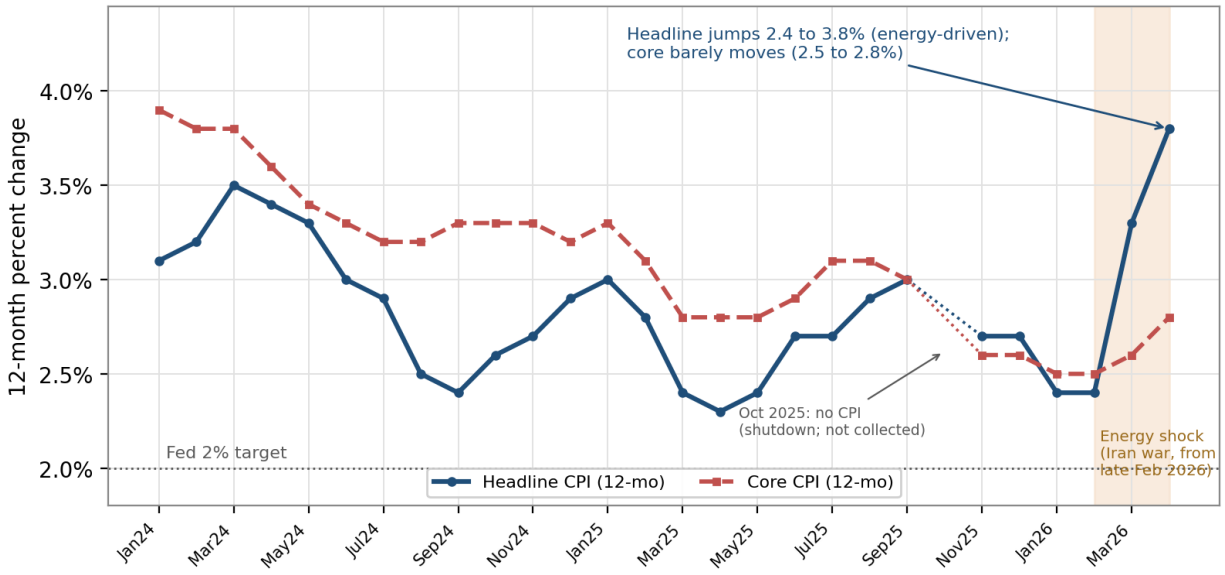
5. Empirical baseline

We anchor the diagnosis in the data as it currently stands, and we are deliberate about what the data does and does not yet show. The diagnosis makes an unusual empirical prediction for this stage: that the headline aggregates will look unremarkable while the fragmentation forms beneath them. The baseline below is consistent with that prediction. The one large movement in the aggregate price level over the recent window is a commodity shock, not a structural signal, and the clearest evidence of fragmentation is found not in the price indices but in the migration of settlement outside the measured monetary system.

5.1 The aggregate is quiet, and what moves in it is energy

U.S. consumer price inflation eased to 2.4 percent over the twelve months to January and February 2026, within reach of the target. It then jumped to 3.3 percent in March and 3.8 percent in April, the highest reading since May 2023. The cause is identifiable and external. Following the escalation of conflict involving Iran in late February 2026, energy prices rose sharply: the energy index was up 17.9 percent over the year to April and gasoline 28.4 percent, and energy accounted for more than two-fifths of the headline gain. Core inflation, which excludes food and energy, moved only from 2.5 to 2.8 percent over the same months. Figure 1 shows the two series.

Figure 1. Headline vs. core CPI: the 2026 spike is an energy shock, not structural



Source: U.S. Bureau of Labor Statistics, CPI-U, all items and all items less food and energy, 12-month change, NSA.
BLS did not collect survey data for October 2025 (appropriations lapse); the dotted segment bridges Sept to Nov 2025.

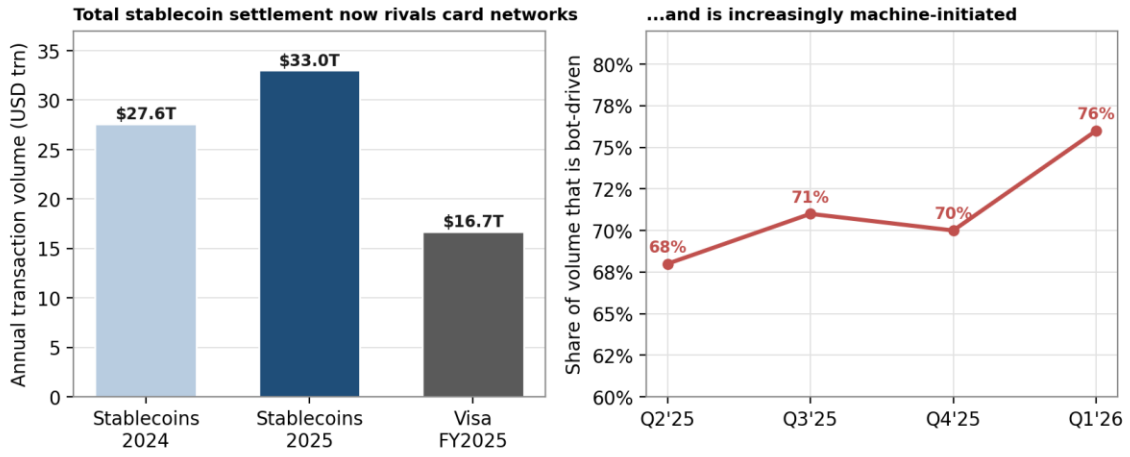
Two points follow. First, the single large movement in the aggregate over this window is a commodity-price shock, and the widening gap between headline and core is its signature, not the signature of the data-money fragmentation. Second, and for our purposes more important, the aggregate is not currently registering the fragmentation at all. This is exactly what the diagnosis predicts. A bifurcation between authenticated and synthesizable goods

operates within the core, between categories, and a fixed-weight aggregate that averages over them will look quiet even as the underlying dispersion grows. The baseline therefore offers no comfort to an aggregate-first reading: the absence of an AI signal in the headline is consistent with the fragmentation being real and early, not with its being absent.

5.2 Value migrating outside the measured system

The clearest contemporary evidence sits on the monetary-fragmentation side, in the migration of settlement onto rails the conventional monetary aggregates do not capture. Stablecoin transaction volume reached roughly 33 trillion dollars in 2025, an increase of about 72 percent over the 27.6 trillion recorded in 2024, placing annual stablecoin settlement on the order of the major card networks, against a circulating supply near 320 billion dollars in mid-2026. More telling than the level is its composition: the share of settlement volume that is machine-initiated rather than retail rose from roughly 68 percent in the second quarter of 2025 to about 76 percent in the first quarter of 2026. Figure 2 shows both facts.

Figure 2. A measurable symptom of monetary fragmentation: value settling outside the measured aggregates



Sources: Artemis Analytics / Visa annual reporting / CEIC or equivalent transaction methodology notes, latest available estimates. Stablecoin figures are raw total transaction volume and include automated, repeated, and potentially non-final transactions. Adjusted estimates of final economic activity are materially lower. Visa figure refers to reported payment volume or processed volume, depending on source definition. The comparison is intended to show scale and migration toward machine-legible settlement rails, not a one-for-one comparison with consumer card payments.

We read this as an early and measurable symptom of the monetary-fragmentation side of the wedge, not as its source. Value is increasingly settled on rails outside the standard aggregates, and increasingly between automated agents rather than people, which is the monetary counterpart of an economy whose production has migrated toward machine-legible channels. One qualification matters: the 33 trillion is unadjusted total volume, and adjusted estimates of economic, non-automated activity are considerably lower, on the order of 10 trillion by one method and 26 trillion by another. Even on the conservative estimate, the scale and the rising machine-initiated share are consistent with settlement migrating beyond the perimeter the aggregates were designed to measure. The raw volume figure is not equivalent to final economic payment volume. It includes automated, repeated, and potentially non-final transactions. For that reason, the claim should be read as evidence that settlement activity is migrating toward machine-legible rails, not as a one-for-one comparison with card-network consumer payments.

5.3 What the baseline cannot yet show

The data-side symptom, the provenance premium on verified-human data and the conversion gap between the value individuals create and the money they receive, has no clean published series, and constructing one is the empirical heart of the agenda set out in Section 9. This is a limitation of the present evidence and, at the same time, a prediction of the diagnosis. The fragmentation is early, and the measurement apparatus is built to miss it, which is precisely why the falsification conditions of Section 3.6 are framed as divergence patterns to be tracked rather than as a level to be read off an existing index. The baseline establishes the two things the diagnosis needs at this stage: the aggregate is quiet, and value has begun to move outside it.

6. Four transmission channels

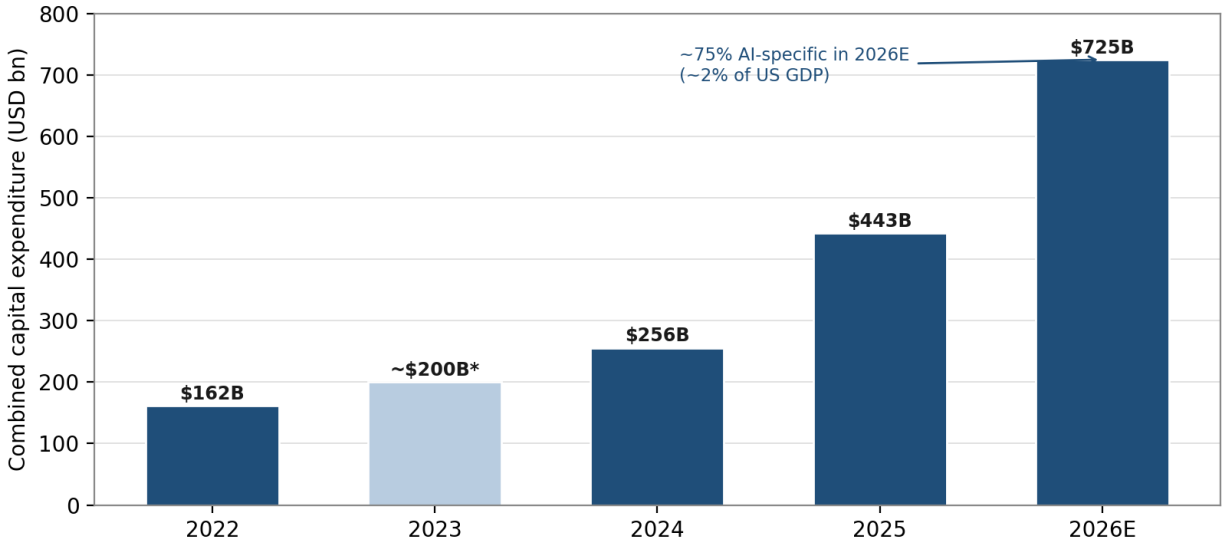
The fragmentation propagates into observable macroeconomic effects through four channels. Each is a consequence of the wedge rather than an independent mechanism. We present four, having folded an earlier sovereign-financing channel into the first.

6.1 Channel 1: capex and the Cantillon effect

The construction of AI capacity injects spending into the economy at identifiable first-receiver points before prices and wages fully adjust, the classic Cantillon configuration in which the distribution of a monetary injection depends on where it lands first. The scale is now macroeconomic. Combined capital expenditure by the five largest hyperscalers rose from about 162 billion dollars in 2022 to 256 billion in 2024 and 443 billion in 2025, and first-quarter 2026 earnings put 2026 on track for roughly 725 billion, of which market estimates suggest about three-quarters is AI-specific, an amount on the order of two percent of GDP. Figure 3 shows the trajectory. The 2025 figures are reported actuals where available;

the 2026 figure is an annualized estimate; and the AI-specific share is estimated rather than reported.

Figure 3. The Cantillon first receivers: Big-5 hyperscaler capex (Channel 1)



Sources: Company filings and earnings releases for Alphabet, Amazon, Meta, Microsoft, and Oracle; 2026 figure is an annualized estimate based on latest available quarterly run rates; AI-specific share is a market estimate, not a reported aggregate. Figures are rounded and should be read as order-of-magnitude evidence of first-receiver spending, not as a calibrated macro series.

The first receivers of this spending are the providers of compute and infrastructure and their supply chains, together with the equity holders who finance them, rather than the financial intermediaries through which conventional monetary injections pass. That difference makes AI capex a distinctive test of a contested question. The literature disagrees about the distributional sign of large monetary injections: euro-area evidence finds that expansion compressed income inequality through employment while leaving wealth inequality little changed (Lenza & Slacalek, 2024), whereas Danish administrative evidence finds that easing raised income inequality through non-labor channels (Andersen et al., 2023). An injection whose first receivers are the owners of data and compute is, on its face, more likely to follow the second pattern than the first, and tracking its incidence is a concrete empirical task the channel defines. Where AI infrastructure is financed or guaranteed through public vehicles,

the first-receiver effect is amplified, which is the sovereign-financing mechanism we fold into this channel.

6.2 Channel 2: energy and compute competition

AI computation competes for energy and grid capacity with other uses, placing upward pressure on the relative price of energy-intensive goods and of compute itself. This channel must be distinguished carefully from the commodity shock of Section 5.1: the energy spike of early 2026 was geopolitical in origin, whereas the channel here is the structural competition for power that large-scale training and inference impose on a constrained grid. We flag that the academic grounding for this channel is thinner than for the other three, since the available evidence is largely industry and agency data rather than peer-reviewed estimates, and we list its formalization as a priority in Section 9. It is included because the mechanism is real and first-order for the relative-price structure even where the magnitude is not yet pinned down.

6.3 Channel 3: cognitive-labor repricing

As AI substitutes for cognitive tasks, the task-based framework predicts a repricing of the affected labor (Acemoglu & Restrepo, 2018, 2022). This is the channel through which the displacement of human cognitive work, the data side's substitution of synthesizable output for human output, shows up as wage and employment effects, and it is the channel most directly continuous with the existing automation literature. In the language of Section 4 it is the force that lowers the labor cost in the synthesizable sector and thereby drives down the price of synthesizable goods, one half of the bifurcation. Its incidence falls on the workers whose tasks are most readily synthesized, which need not be the low-wage workers on

whom earlier waves of automation fell, a redistribution within labor that the aggregate wage bill conceals.

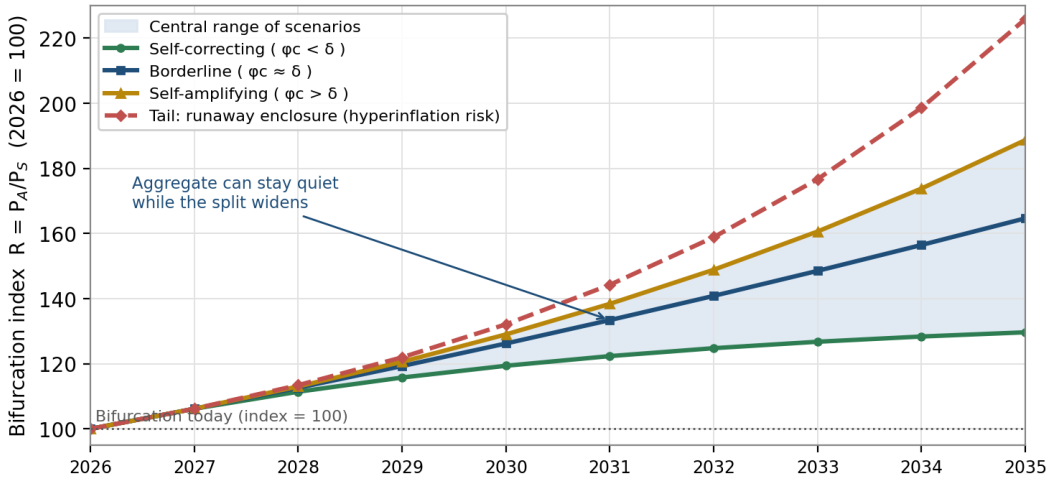
6.4 Channel 4: heterogeneous incidence

Because automation raises the return to the capital that does the automating as well as displacing labor, it widens both income and wealth inequality (Moll, Rachel, & Restrepo, 2022). The wedge sharpens this. The surplus created at the data-money interface accrues to the controllers of data and compute, the ownership axis of Section 3.3, a margin that aggregate factor-share accounting flattens because it does not distinguish households by their access to the conversion interface. This is the channel that requires a heterogeneous-agent treatment, sketched in the appendix, and it is where the model's two distributional results bite: the cross-sectional variance of household inflation scales with the squared sectoral inflation gap (Proposition 2), and a single mean-targeting instrument does not generally stabilize that variance if it leaves the bifurcation gap unchanged (Proposition 5). Channel 4 is, in effect, the others viewed through the distribution rather than the aggregate.

7. Scenario analysis

Rather than forecast a single path, we follow the scenario method of Korinek and Suh (2024) and describe a bounded fan of trajectories that differ in a small number of parameters: the speed of cognitive-task displacement, the elasticity of the energy supply response, the fiscal posture toward the first receivers of AI capex, and the speed at which value settlement migrates outside the measured system. In the model of Section 4 these map onto a single object, the loop gain of the reinforcing feedback, and the regimes of Proposition 4 organize the fan. Figure 4 shows the resulting paths for the bifurcation. It is a model-based illustration, not a forecast, and its magnitudes are not calibrated.

Figure 4. Scenario paths for the bifurcation, by feedback regime (illustrative, model-based)



Model-based projection, not a forecast. Paths simulate the enclosure recursion of equations (10)-(11), Section 4, from a common base under three regimes of the loop gain λ = 1 + φc − δ (self-correcting, borderline, self-amplifying), plus the self-amplifying tail, mapped to the bifurcation R = P_A/P_S. Parameters are illustrative; the figure shows qualitative regimes implied by Proposition 4, not calibrated magnitudes.

In the *self-correcting* regime, where competition and regulation erode enclosure faster than the premium on controlling authenticated data rebuilds it, the bifurcation widens and then levels off; the fragmentation is real but bounded. In the *borderline* regime the bifurcation widens steadily without accelerating, and this is the path most consistent with the stagnationist reading of Acemoglu (2025): aggregate productivity gains are modest, the headline price level stays quiet, and yet the distribution of prices and incomes pulls apart underneath. In the *self-amplifying* regime the loop gain exceeds the rate at which enclosure is eroded, and the bifurcation accelerates until it becomes visible even in the aggregates.

Beyond these lies a tail. If enclosure runs away, monetary fragmentation can accelerate to the point where value migrates out of the unified system fast enough that singleness erodes in practice, conventional anchors lose traction, and the price path of authenticated goods enters an unstable acceleration. This is the only scenario in which the language of hyperinflation is warranted, and we are deliberate in placing it where it belongs: as a tail conditional on the conversion interface being left unrepaired, not as a central case. This is

not a forecast and not the central scenario. Naming it is part of the diagnosis, because the feedback structure of Section 4 is what makes such a tail possible; treating it as the headline would misrepresent the analysis. The central scenario remains fragmentation and bifurcation beneath quiet aggregates, not hyperinflation.

8. Implications for monetary policy

If the diagnosis holds, conventional monetary policy faces a dimensional problem rather than merely a calibration problem. The standard toolkit sets one policy rate with reference to one aggregate price index, on the presumption of a unified substrate. When the substrate is fragmenting, so that the price level is bifurcating and value is migrating outside the measured aggregates, a single instrument referenced to a single index may be insufficient if its operative channel moves the mean while leaving the bifurcation gap unchanged. In the reduced-form case considered above, tightening that restrains the authenticated-goods price path may leave the synthesizable basket on a different trajectory; the average can be stabilized while the split remains or widens.

Proposition 5 makes this precise. The welfare-relevant object is the distribution of household inflation rates, whose loss decomposes into a level term and a dispersion term. Within this reduced-form setup, a single instrument that moves only the mean can stabilize the average but does not generally stabilize the dispersion, so the minimized loss is an irreducible residual that grows with the bifurcation. Conventional single-index policy is insufficient if its operative channel moves the mean price level but leaves the bifurcation gap unchanged.

This extends the heterogeneous-agent insight that aggregate instruments transmit unevenly across a distribution (Kaplan, Moll, & Violante, 2018; Auclert, 2019). That literature shows the transmission of policy is distributionally uneven within a unified system. The fragmentation carries the point one step earlier, to the measurement on which policy is set: the central bank is steering by an index that, by construction, averages away the phenomenon it would need to observe. The measurement problem precedes the instrument problem, since an authority that cannot see the bifurcation cannot respond to it whatever instruments it holds.

We do not propose a new instrument or institution here. The design of a policy response, whether through provenance-aware measurement, a second instrument aimed at the dispersion, or arrangements that repair the conversion interface directly, is the institutional question the team has deliberately separated from this paper. What the diagnosis establishes is the prerequisite: that the measurement of the bifurcation and of out-of-system settlement is the precondition for any response, and is therefore a research priority in its own right. This is a diagnosis of inadequacy, not a policy proposal.

9. Open questions and research agenda

The diagnosis implies a program. We group the open questions into four.

Formalizing the coupling. The feedback loop of Section 3.5, written locally in Section 4, should be developed into a fully solved two-substrate model in which a non-singular data input and a singular monetary unit interact and the bifurcated price level emerges endogenously. The central question the model must resolve is which side of the loop leads: whether fragmentation of the data input is the driving margin and the price-level bifurcation

the response, or whether monetary fragmentation, in the form of settlement and value-storage arrangements proliferating outside the unified system, is the leading edge that pulls data into enclosure. Establishing this temporal leadership requires additional structure on shocks, timing, and information flow, together with the global dynamics beyond the local linearization of Proposition 4.

A heterogeneous-agent extension. The incidence channel, Channel 4 of Section 6, calls for an extension of the heterogeneous-agent New Keynesian framework (Kaplan, Moll, & Violante, 2018; Auclert, 2019) in which households differ in the provenance composition of their consumption baskets and in their access to the conversion interface. The appendix sketches the structure; calibrating it is deferred, and the sketch is conceptual rather than a calibrated result.

Measurement. The falsification conditions of Section 3.6 require data that does not yet exist in convenient form: a price index decomposed by provenance, separating authenticated from synthesizable baskets; a measure of value settled outside the standard monetary aggregates; and a measure of the gap between the value individuals create through data production and the money they receive for it. The calibration of the model's fragmentation wedge and feedback sensitivities depends on the same series. Constructing them is the empirical heart of the agenda and the natural bridge to a subsequent empirical paper.

Channel-level tests. Each transmission channel defines a concrete empirical question: the incidence of AI-capex injections by first-receiver position (Channel 1), the structural relationship between compute demand and energy prices net of commodity shocks (Channel 2), and the distribution across workers of cognitive-task displacement (Channel 3). These

are estimable with existing methods once the relevant series are assembled, and they are the most immediate point of contact between the diagnosis and conventional empirical work.

Comparative regime test: China, the United States, and Europe. The model predicts that data-governance regimes should shape the wedge. A China and rest-of-world comparison would test whether stronger localization, state-guided AI deployment, and distinct payment and settlement infrastructures alter the relationship between data enclosure, provenance premiums, and household-level price incidence. This test would move the diagnosis from a general AI-macro claim to a comparative institutional research agenda.

Table 1 sets the program out as a set of falsifiable hypotheses, each with an observable measure, a candidate data source, a concrete falsification test, and a comparative dimension across the China, United States, and Europe regimes.

Table 1. A research design for the data–money wedge.

Hypothesis	Observable measure	Candidate data source	Falsification test	China / US / EU comparison
Provenance premium rises	Licensing terms and certification fees for human-authored versus model-generated corpora	Data-licensing disclosures; provenance-standard registries	Weakens if the premium on verified human data does not rise as the synthetic-content share grows	Compare premia under China's state-guided pooling, US platform licensing, and EU consent rules
Conversion wedge is large (low pass-through $\bar{\tau}$)	Data-derived revenue booked by controllers versus household data income	Firm filings; national accounts; household income surveys	Weakens if household data income rises toward the data value captured by controllers	Compare household pass-through across ownership and compensation regimes
Authenticated versus synthesizable price gap widens (R)	Provenance-decomposed price indices for the two baskets	Transaction or scanner data tagged by provenance; constructed sub-indices	Weakens if the two baskets do not diverge during periods of rising AI adoption and data enclosure	Compare divergence across regimes with different authentication rules
Household inflation dispersion rises (Prop. 2)	Cross-sectional variance of household inflation, by basket provenance share	Household-level price and expenditure microdata	Weakens if dispersion does not rise with the authenticated-synthesizable gap	Compare dispersion dynamics across regimes
Settlement migrates outside conventional aggregates	Share of settlement outside standard aggregates; machine-initiated share	On-chain settlement data; payment-network and central-bank statistics	Weakens if out-of-aggregate settlement does not grow as AI activity rises	Compare migration under distinct payment and settlement infrastructures
AI-capex first-receiver incidence	Distribution of AI-capex income by recipient position	Firm filings; ownership and capital-flow data	Weakens if AI-capex income does not concentrate among first receivers relative to broad labor	Compare incidence under state-financed versus market-financed buildout

Hypothesis	Observable measure	Candidate data source	Falsification test	China / US / EU comparison
Policy-rank condition (Prop. 5)	Co-movement of policy rate, mean inflation, and inflation dispersion	Policy-rate series; inflation-dispersion estimates	Weakens if single-index policy is seen to stabilize dispersion as well as the mean	Compare policy frameworks and dispersion outcomes across regimes

The hypothesis does not require, and the paper does not evaluate, any specific private implementation.

10. Conclusion

The debate over AI and the macroeconomy has been a debate about the size of an aggregate. We have argued that the aggregate is the wrong object. The digital and the real economy are one coupled system, and that system is fragmenting at the interface where data produced by individuals is converted into money, or fails to be. The wedge is the joint product of a data fragmentation, in which data is a non-singular input that cannot clear at par across jurisdiction, platform, sector, provenance, and ownership, and a monetary fragmentation that surfaces as a bifurcated price level. The two reinforce one another, and the standard instruments are built to miss the result.

We have given that diagnosis a formal skeleton, in which the bifurcation, the heterogeneous incidence of inflation, the uncompensated conversion wedge, and the dimensional inadequacy of single-instrument policy follow as formal results; we have anchored it in the data as it stands, where the aggregate is quiet and value has begun to migrate outside it; and we have specified the falsifiable claim and the program that would test it. Whether the aggregate productivity gain from AI proves large or modest, the fragmentation is the consequential phenomenon, and specifying how to prove it is the work this paper sets out. We make no policy proposal here; the diagnosis comes first.

Disclosure statement

The authors have professional affiliations with DrumWave and with the International Data Reserve (IDR), organizations whose activities relate to the data-economy questions discussed in this paper. The analytical claims advanced here are not contingent on the success of any specific commercial architecture.

References

Acemoglu, D. (2025). The simple macroeconomics of AI. Economic Policy, 40(121), 13–58. https://doi.org/10.1093/epolic/eiae042

Acemoglu, D., & Restrepo, P. (2018). The race between man and machine: Implications of technology for growth, factor shares, and employment. American Economic Review, 108(6), 1488–1542. https://doi.org/10.1257/aer.20160696

Acemoglu, D., & Restrepo, P. (2022). Tasks, automation, and the rise in U.S. wage inequality. Econometrica, 90(5), 1973–2016. https://doi.org/10.3982/ECTA19815

Acquisti, A., Taylor, C., & Wagman, L. (2016). The economics of privacy. Journal of Economic Literature, 54(2), 442–492. https://doi.org/10.1257/jel.54.2.442

Agrawal, A., Gans, J., & Goldfarb, A. (Eds.). (2019). The economics of artificial intelligence: An agenda. University of Chicago Press / NBER. https://doi.org/10.7208/chicago/9780226613475.001.0001

Andersen, A. L., Johannesen, N., Jørgensen, M., & Peydró, J.-L. (2023). Monetary policy and inequality. Journal of Finance, 78(5), 2945–2989. https://doi.org/10.1111/jofi.13262

Argente, D., & Lee, M. (2021). Cost of living inequality during the Great Recession. Journal of the European Economic Association, 19(2), 913–952. https://doi.org/10.1093/jeea/jvaa018

Arrieta-Ibarra, I., Goff, L., Jiménez-Hernández, D., Lanier, J., & Weyl, E. G. (2018). Should we treat data as labor? Moving beyond "free." AEA Papers and Proceedings, 108, 38–42. https://doi.org/10.1257/pandp.20181003

Auclert, A. (2019). Monetary policy and the redistribution channel. American Economic Review, 109(6), 2333–2367. https://doi.org/10.1257/aer.20160137

Bank for International Settlements. (2025). The next-generation monetary and financial system (Annual Economic Report 2025, Ch. III). https://www.bis.org/publ/arpdf/ar2025e3.htm

Bergemann, D., Bonatti, A., & Gan, T. (2022). The economics of social data. RAND Journal of Economics, 53(2), 263–296. https://doi.org/10.1111/1756-2171.12407

Brynjolfsson, E. (2022). The Turing trap: The promise and peril of human-like artificial intelligence. Daedalus, 151(2), 272–287. https://doi.org/10.1162/daed_a_01915

Cory, N., & Dascoli, L. (2021). How barriers to cross-border data flows are spreading globally, what they cost, and how to address them. Information Technology and Innovation Foundation (ITIF).

Farboodi, M., & Veldkamp, L. (2021). A growth model of the data economy. NBER Working Paper No. 28427. https://doi.org/10.3386/w28427

Gorton, G. B., & Zhang, J. Y. (2023). Taming wildcat stablecoins. University of Chicago Law Review, 90(3), 909–971.

Jaravel, X. (2019). The unequal gains from product innovations: Evidence from the U.S. retail sector. Quarterly Journal of Economics, 134(2), 715–783. https://doi.org/10.1093/qje/qjy031

Jaravel, X. (2021). Inflation inequality: Measurement, causes, and policy implications. Annual Review of Economics, 13, 599–629. https://doi.org/10.1146/annurev-economics-091520-082042

Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the economics of data. American Economic Review, 110(9), 2819–2858. https://doi.org/10.1257/aer.20191330

Kaplan, G., Moll, B., & Violante, G. L. (2018). Monetary policy according to HANK. American Economic Review, 108(3), 697–743. https://doi.org/10.1257/aer.20160042

Kaplan, G., & Schulhofer-Wohl, S. (2017). Inflation at the household level. Journal of Monetary Economics, 91, 19–38. https://doi.org/10.1016/j.jmoneco.2017.08.002

Korinek, A., & Suh, D. (2024). Scenarios for the transition to AGI. NBER Working Paper No. 32255. https://doi.org/10.3386/w32255

Lenza, M., & Slacalek, J. (2024). How does monetary policy affect income and wealth inequality? Evidence from quantitative easing in the euro area. Journal of Applied Econometrics, 39(5), 746–765. https://doi.org/10.1002/jae.3053

Moll, B., Rachel, L., & Restrepo, P. (2022). Uneven growth: Automation's impact on income and wealth inequality. Econometrica, 90(6), 2645–2683. https://doi.org/10.3982/ECTA19417

Posner, E. A., & Weyl, E. G. (2018). Radical markets: Uprooting capitalism and democracy for a just society. Princeton University Press.

Reddit, Inc. (2024). Registration statement on Form S-1. U.S. Securities and Exchange Commission.

Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. Nature, 631(8022), 755–759. https://doi.org/10.1038/s41586-024-07566-y

Veldkamp, L., & Chung, C. (2024). Data and the aggregate economy. Journal of Economic Literature, 62(2), 458–484. https://doi.org/10.1257/jel.20221580

Appendix A. A heterogeneous-agent sketch (uncalibrated)

This appendix presents a conceptual sketch, not a calibrated model and not a result. It indicates what formalizing the incidence channel of Section 6.4 would require; the full treatment is deferred to future work (Section 9).

Consider a heterogeneous-agent New Keynesian environment in the tradition of Kaplan, Moll, and Violante (2018) and Auclert (2019), extended in two ways. First, the consumption basket is decomposed into an authenticated component, whose relative price rises with the provenance premium, and a synthesizable component, whose relative price falls as AI lowers its production cost, with households differing in the weight of each in their basket, the parameter ω_i of Section 4. Second, households differ in their access to the conversion interface, the parameter τ_i, that is, in the degree to which the value of the data they produce is realized as income rather than accruing to a controlling firm.

In such an environment a single aggregate price index is a weighted average over heterogeneous baskets and conceals the bifurcation, and a single policy rate transmits asymmetrically across households sorted by basket composition and by interface access. The objects the model would deliver are the distribution of provenance-weighted inflation across households, the incidence of AI-capex injections by first-receiver position, and the conditions on the feedback of Section 4.5 under which the provenance premium is self-amplifying. Calibration would require the provenance-decomposed price series and the out-of-system settlement measure identified in Section 9. We state the structure and stop.