Artificial intelligence and monetary policy: A framework and perspective on cyclical transmission, structural transition, and financial stability

Simone Lenzu

Federal Reserve Bank of New York
New York University
CEPR

February, 2026

Abstract

I develop a framework to analyze how the diffusion of artificial intelligence (AI) can affect monetary policy through three interrelated channels: cyclical transmission, structural transition, and financial stability. In the short run, AI can reshape inflation dynamics by altering how supply and demand disturbances transmit into prices—through changes in firms' pricing behavior, production technologies, cost pass-through, and expectations formation—even when conventional measures of economic slack are unchanged. Over longer horizons, AI may induce a structural transition by affecting productivity, investment, and risk-taking, thereby shifting key equilibrium benchmarks around which monetary policy is calibrated, including potential output and the natural rate of interest. Finally, AI presents both opportunities and risks for financial stability: while it may improve information processing, credit allocation, and financial inclusion, it can also foster model monocultures and amplify expectations-driven asset valuation dynamics, increasing the likelihood of financial distress. I argue that AI does not call for a redefinition of central banks' objectives, but it does require a more nuanced application of existing frameworks, as its rapid diffusion complicates blurs the distinction between cyclical dynamics and structural shifts in economic fundamentals.

Email: simone.lenzu@ny.frb.org and slenzu@stern.nyu.edu.

I am grateful to Nicola Cetorelli, Marco Del Negro, Keshav Dogra, Fulvia Fringuellotti, Danial Lashkari, Paolo Pesenti, Giorgio Topa, and John Williams for useful feedback, suggestions, and discussions. The views expressed in this paper are those of the author and do not necessarily reflect those of the Federal Reserve Bank of New York and the Federal Reserve System, or any other institution with which the author is affiliated. All mistakes are my own.

1 Introduction

Artificial intelligence (AI) is rapidly emerging as a transformative technology with the potential to reshape many, if not most, aspects of economic activity, including how goods are produced, how workers contribute to production, how prices are set, how expectations are formed, and how risks are assessed. Given that central bank mandates center on price stability and financial stability, these developments place AI squarely within the domain of central banking. Indeed, there is a growing consensus that the relevant question is no longer whether the diffusion of AI will matter for monetary policy, but how. A broad-based adoption of AI is likely to alter the behavior of the variables that monetary policy is designed to stabilize—namely inflation dynamics, real economic activity relative to potential, and financial conditions—and, in doing so, may require policymakers to adapt how they interpret and respond to business-cycle fluctuations.

In this paper, we offer a perspective on these issues with a specific focus on the recent wave of advances in artificial intelligence. Although we refer broadly to AI, the analysis is motivated in particular by the rapid development and diffusion of large-scale, general-purpose, and generative AI (GenAI) models. Unlike earlier, more task-specific forms of automation, these technologies can be flexibly applied across a wide range of activities, augment existing AI tools, and be integrated deeply into core production, decision-making, and organizational processes (e.g., Acemoglu and Restrepo 2020; Brynjolfsson et al. 2021; Aghion et al. 2019; Brynjolfsson et al. 2023).[1] It is this breadth and depth that motivates treating the diffusion of AI as a development of direct relevance for central banks—not because of its technological novelty per se, but because of its potential to alter policy conduct and the risks that central banks seek to manage.

Guided by a stylized theoretical framework, we organize the discussion around three conceptually distinct but interrelated sets of channels through which AI may affect central banks' policies: the short-run transmission of supply and demand shocks into inflation, the long-run transition in economic fundamentals and equilibrium benchmarks, and the implications for financial markets and financial stability.

[1] By lowering the cost of information processing, prediction, and content generation, recent AI innovations increasingly resemble a general-purpose technology with the potential to reshape how goods and services are produced, how tasks are organized within firms, and how firms scale and compete. Emerging evidence suggests that generative AI, in particular, can augment worker productivity across a broad set of tasks rather than simply automating narrow activities, reinforcing its economy-wide scope.

In the short run, taking equilibrium benchmarks and long-run fundamentals as given, AI can affect the **cyclical transmission** of supply and demand disturbances into inflation. On the supply side, AI reshapes how cyclical fluctuations translate into inflation by altering both the mapping from economic slack to real marginal costs and the pass-through of marginal-cost movements into prices. These effects operate through changes in scale and utilization margins, input-market dynamics, and efficiency wedges, as well as through shifts in nominal rigidities, strategic complementarities in price setting, and macroeconomic complementarities in production. On the demand side, AI can influence expectations about future productivity, income, and profitability, affecting current expenditure and inflation dynamics even before productivity gains are fully realized. Together, these channels determine how real shocks propagate into inflation over the business cycle and form the focus of Section 2.

In the long run, by reshaping the economy's underlying fundamentals, AI is likely to be a key force behind a **structural transition** that affects both long-run growth trends and the equilibrium benchmarks around which monetary policy stabilizes the economy. Through its impact on technological change, investment opportunities, risk, and market structure, AI influences potential output and the natural rate of interest, as well as the intertemporal elasticity of substitution that governs the sensitivity of aggregate demand to real interest rates. These effects operate at low frequency and define the evolving benchmarks around which monetary policy must be calibrated, rather than shaping the short-run propagation of shocks. We discuss these potential long-run effects in Section 3.

Finally, AI may have a direct bearing on the conduct of central banking through its impact on the financial system and on **financial stability**. As discussed in Section 4, the diffusion of AI technologies is reshaping key segments of the financial system—including payments, lending, insurance, and asset management—by altering information processing, risk-taking behavior, and the structure of financial intermediation. These changes also affect the transmission of monetary policy, as shifts in the structure of intermediation and risk-taking alter how policy rates map into financial conditions faced by households and firms. At the same time, financial stability concerns are amplified by the possibility that expectations about AI-driven productivity gains may prove overly optimistic or materialize only gradually. In such an environment, elevated asset valuations, leverage, and reliance on nonbank financing can interact with expectations-driven dynamics to generate systemic vulnerabilities, even in the absence of traditional macroeconomic imbalances.

Section 5 places the analysis in a policy perspective and concludes. I argue that artificial intelligence does not call for a redefinition of monetary policy objectives, but it does change the environment in which those objectives are pursued. The diffusion of AI may weaken the informational content of traditional indicators such as the unemployment gap or the output gap, and it may alter—while also increasing uncertainty about—the trade-offs between inflation and economic activity. This raises the value of robust policy design and greater reliance on cost-side diagnostics and pricing behavior, rather than exclusive dependence on reduced-form Phillips curve estimates.

A related challenge is that AI-driven structural change can blur the distinction between cyclical fluctuations and shifts in equilibrium benchmarks such as potential output or the natural rate of interest. Transitional frictions and reorganization costs may temporarily depress effective productivity even as the technological frontier expands, complicating the interpretation of inflationary pressures. At the same time, faster information flows and more responsive expectations may make policy both more powerful and more fragile, as errors in assessing costs, benchmarks, or risks propagate more rapidly through prices, expectations, and firms', households', and intermediaries' balance sheets.

2 AI and the transmission of cyclical disturbances

We use a representative-agent New Keynesian framework to study how AI affects the short-run transmission of cyclical supply and demand disturbances into inflation. Households choose consumption and saving subject to an intertemporal Euler equation. Firms operate under imperfect competition, and prices are subject to nominal and real rigidities that prevent costless and instantaneous reoptimization. Factor markets clear, but may be subject to frictions that drive a wedge between factor prices and marginal products.

We use lowercase letters to denote natural logarithms of the corresponding variables; starred variables denote natural (flexible-price) benchmarks that anchor the cyclical model—such as potential output, the natural wage, and the natural rate of interest; and hats denote log deviations from those benchmarks (e.g., $\widehat{y}_t \equiv y_t - y_t^\star$ denotes the output gap). For the purpose of analyzing how AI affects inflation dynamics in response to business-cycle shocks, these equilibrium benchmarks are taken as given. In the long run, however, such natural benchmarks are themselves endogenous outcomes of technological, institutional, and market-structure forces, which may be reshaped by the diffusion of AI technologies—a point we return to in Section 3.

In this framework, aggregate inflation is determined by the interaction of two blocks. The first is an aggregate supply block–the Phillips curve–which links inflation to a measure of economic slack through firms' price-setting decisions. The second is an aggregate demand block–the investment-saving (IS) relation–which links aggregate expenditure to the real interest rate through households' intertemporal substitution. In the short run, AI matters insofar as it alters the transmission of shocks through these two blocks: by changing how movements in marginal costs are transmitted into inflation via the Phillips curve, and by shaping the propagation of demand shocks through expectations in the IS equation.

2.1 Cyclical transmission through the supply side

The New Keynesian Phillips curve (NKPC) lies at the core of modern analyses of short-run inflation dynamics. In its standard formulation, the NKPC implies that inflation depends on expected future inflation and on measures of real economic activity relative to their potential levels, as determined by underlying fundamentals. we build on Gagliardone et al. 2025b and spell out a (generalized) cost-based formulation of the NKPC that allows for feedbacks between inflation, input-market frictions, and slow-moving technological change.

In its primitive formulation, the NKPC is a forward-looking difference equation relating inflation to expected future inflation and the real marginal cost gap, defined as the log deviation of real marginal cost from its flexible-price level, $\widehat{mc}_t := mc_t - mc_t^{\star}$:

$$\begin{aligned}\pi_t &= \lambda\, \widehat{mc}_t + \beta\, \mathbb{E}_t\{\pi_{t+1}\} \\ &= \lambda \left(\widehat{q}_t - \widehat{a}_t\right) + \beta\, \mathbb{E}_t\{\pi_{t+1}\} \qquad (1)\end{aligned}$$

The coefficient λ denotes the slope of the cost-based NKPC. Intuitively, movements in $\widehat{mc}_t$ capture the extent to which cyclical demand- and supply-side disturbances strain or relieve productive resources relevant for firms' pricing decisions; the slope λ governs how strongly these cost fluctuations are transmitted into inflation. Due to nominal rigidities, firms' pricing decisions are forward-looking: prices set today depend not only on current marginal costs but also on anticipated future economic conditions. This gives rise to the expectational term $\mathbb{E}_t\{\pi_{t+1}\}$. The parameter $\beta \in (0,1)$, reflecting firms' effective discounting of future profits, captures the degree of forward-lookingness in price setting. It is useful to further decompose $\widehat{mc}_t$ into two cyclical components: a *cost pressure index* $\widehat{q}_t$—capturing scale effects and direct input pressure associated with expanding production—and the *effective productivity* of the economy, $\widehat{a}_t$. (We formally define both quantities and their components below.) The intuition is straightforward. Improvements in technology, organization, or input efficiency reduce the amount of resources required to produce a given level of output, thereby lowering marginal costs. Inflationary pressure emerges when unit production costs rise faster than productivity; conversely, productivity gains can offset cost pressures and dampen inflation even during expansions.

Equation (1) highlights that the short-run transmission of real economic fluctuations into inflation operates through two distinct margins: (*i*) a real-side channel, governing how cyclical disturbances translate into fluctuations in real marginal costs; and (*ii*) a pass-through channel, governing how a given cost shock is transmitted into inflation through firms' pricing decisions via the slope λ. The diffusion of AI can reshape the transmission of shocks into inflation through both channels, even when the underlying equilibrium benchmarks remain unchanged. We discuss them in turn.

Effects of AI on cyclical moments of real marginal cost

Under rather general assumptions, up to first order, we can further decompose the cyclical deviations of real marginal costs from trend into variation in real unit cost into pressures arising from scale effects associated with expanding production (the output gap $\widehat{y}_t$), tighter factor markets (the real user cost index gap, $\widehat{w}_t$), and cyclical wedges in factor markets ($\widehat{\tau}_t$):

$$\widehat{mc}_t \;=\; \widehat{q}_t - \widehat{a}_t \;=\; \left(\chi\widehat{y}_t + \widehat{w}_t + \widehat{\tau}_t\right) - \widehat{a}_t.$$

Plugging the expression for $\widehat{mc}_t$ inside the NKPC in (1) we obtain:

$$\pi_t = \kappa\,\widehat{y}_t + \lambda\Big(\widehat{w}_t + \widehat{\tau}_t - \widehat{a}_t\Big) + \beta\,\mathbb{E}_t\{\pi_{t+1}\} \qquad\qquad \kappa := \lambda\chi \tag{2}$$

This decomposition elucidates how different forms of cyclical demand- and supply-side disturbances influence real marginal costs and, conditional on the slope λ, inflation. First, the formulation of the NKPC in (1) nests the conventional formulation of the NKPC, which—under more restrictive assumptions on input markets and preferences—establishes a proportional relationship between the output gap, a standard measure of economic slack, and inflation.[2] The parameter χ captures how strongly deviations of economic activity from potential are transmitted into movements in real marginal cost, as discussed below. Second, this decomposition makes explicit that outside the knife-edge case of frictionless input markets, the output gap is not a sufficient statistic for inflation. In fact, disturbances in input markets and movements in productivity also contribute to shaping firms' cost and therefore their pricing behavior. Singling out these factors is important to gain some perspectives into why and how the diffusion of AI might matter in shaping cost dynamics.

Scale effects—When output rises relative to its natural level—i.e., when the output gap $\widehat{y}_t$ is positive—firms must expand production along increasingly costly margins. The elasticity χ governs the sensitivity of real marginal cost to changes in aggregate output and captures how rapidly production costs rise as activity expands. In the present framework, χ reflects technological scale effects and the availability of short-run adjustment margins, including capacity utilization, inventories, bottlenecks, and the presence of quasi-fixed production factors. A higher χ corresponds to environments in

[2]The derivation of the conventional NKPC, $\pi_t = \tilde{\kappa}\,\widehat{y}_t + \beta\,\mathbb{E}_t\{\pi_{t+1}\}$, relies on a set of convenient but very restrictive assumptions, including separable household preferences over consumption and leisure, competitive labor markets, and flexible wages. Under these conditions, the household intratemporal first-order condition equates the real wage to the marginal rate of substitution between consumption and leisure, allowing real marginal cost to be expressed as a function of aggregate output. Following Galí (2015), we can characterize the slope coefficient $\tilde{\kappa} = \lambda * \tilde{\chi} = \lambda\big(\gamma + \frac{\varphi+\alpha}{1-\alpha}\big)$, where φ denotes the inverse Frisch elasticity of labor supply, α captures the degree of returns to scale in production, and γ the elasticity of substitution across differentiated producers. $\tilde{\chi}$ increases when labor supply is less elastic, when returns to scale are more strongly decreasing, and when the elasticity of substitution across producers is higher. In our generalized NKPC, these forces are captures by the parameter χ as well as the real unit cost gap $\widehat{w}_t$.

Moreover, when the mapping between real marginal cost gap and output gap holds, we can further establish a proportionality between the output gap and the unemployment gap—the deviation of unemployment from its flexible-price (natural) level—yielding an equivalent formulation of the conventional NKPC in terms of labor market tightness: $\pi_t = \tilde{\kappa}^u\,(u_t - u_t^\star) + \beta\,\mathbb{E}_t\{\pi_{t+1}\}$, with $\tilde{\kappa}^u$ proportional to $\tilde{\kappa}$. See Galí (2015) for formal derivations. When labor markets feature frictions or wage rigidities, real marginal costs can log longer be expressed as log-linear mapping of output, the conventional NKPC formulation no longer holds, and inflation dynamics depend instead on wage-setting behavior, giving rise to a wage Phillips curve, as discussed in the text. See, also, Erceg et al. (2000) and Gertler and Trigari (2009).

which scaling up production is costly—for example, due to strongly decreasing returns to scale, tight capacity constraints, limited scope for utilization adjustments, or fragile supply chains. Conversely, a lower χ reflects technologies and organizational structures that allow firms to expand output with relatively small increases in marginal cost by reallocating existing resources, drawing down inventories, or increasing utilization along flexible margins.

The diffusion of AI can materially affect the economy's ability to scale production over the business cycle. By improving demand forecasting, logistics, scheduling, and process optimization, AI can relax effective capacity constraints, enhance inventory management, and expand utilization margins. In such environments, firms can accommodate demand expansions through smoother production schedules, inventory buffers, and more efficient use of existing capital and labor, reducing the sensitivity of marginal costs to output and lowering χ.

Inventory management and production smoothing constitute a particularly important—yet often overlooked—margin of adjustment. AI-enhanced forecasting and dynamic inventory control can decouple sales fluctuations from contemporaneous production, weakening the link between output expansions and marginal cost pressures in the short run. This mechanism effectively flattens the real-side transmission from activity to inflation by lowering χ. By contrast, if AI adoption reinforces just-in-time production, increases reliance on concentrated upstream inputs, or tightens interdependencies across supply chains, inventory buffers may shrink, and marginal costs may respond more sharply to demand-driven expansions, raising χ.

AI can also improve the utilization of existing productive capacity—both capital and labor—through better routing, maintenance, monitoring, and task allocation. To the extent that output can expand along utilization margins rather than through costly factor accumulation, marginal costs rise more slowly with activity. More generally, the presence of utilization, inventories, and other short-run adjustment margins implies that the relevant object for inflation transmission is an effective elasticity χ, shaped by technology, organization, and production planning.

Equilibrium dynamics in factor markets—Cyclical movements in real input prices further contribute to marginal cost dynamics. The term $\widehat{w}_t$ denotes the deviation of the real user cost index of variable inputs—a bundle of labor, intermediates, and to some extent capital user costs—from its natural benchmark.[3] Movements in $\widehat{w}_t$ reflect how equilibrium real factor prices respond to cyclical conditions. In particular, it captures how nominal rigidities, scarcity, adjustment costs, and substitution patterns across inputs translate aggregate fluctuations into movements in real input costs, making variable inputs temporarily more expensive than under flexible-price adjustment.

For our purposes, it is useful to distinguish between two margins that jointly govern the behavior of $\widehat{w}_t$ over the business cycle. The first operates through nominal rigidities and adjustment frictions in factor markets. When wages or other input prices are set in staggered contracts, bargained infrequently, indexed imperfectly, or subject to adjustment costs, real user costs may respond sluggishly or asymmetrically to changes in aggregate demand. Following a demand expansion, firms may wish to expand input usage, but existing contracts or slow price adjustment imply that real input prices do not immediately move to their flexible-price levels. In such environments, variable inputs can become temporarily more expensive than under flexible adjustment, generating a positive $\widehat{w}_t$ even in the absence of changes in technology or factor scarcity.

The second margin operates through the interaction of demand and supply elasticities in factor markets. When aggregate demand rises, firms increase their demand for inputs; the resulting equilibrium response of real input prices depends on the elasticity of factor supply and on substitution possibilities across inputs. If labor supply is inelastic, expanding employment requires relatively large increases in real wages, raising the user cost of labor relative to its natural benchmark. Conversely, when labor supply is elastic, or when firms can substitute toward other inputs or utilization margins, real input prices respond less strongly to cyclical expansions.

AI has the potential to influence inflation dynamics by reshaping both margins simultaneously. By reducing adjustment frictions through improved matching, recruiting, scheduling, and training, AI can increase the speed with which real wages and other input prices respond to changes in demand, dampening cyclical deviations in real input costs. Similar mechanisms apply to intermediate inputs and capital: when their prices adjust sluggishly, demand expansions generate temporary increases in real user costs

[3] As shown in the appendix, the real user-cost index $w_t := w_t^n - p_t$ is a cost-minimizing bundle of the real prices of multiple production inputs—such as different types of labor, capital services, and intermediate goods. Up to a first-order approximation around the natural benchmark, $\widehat{w}_t$ is given by a cost-share-weighted average $\sum_{j=1}^{J} c_j^{\star} \widehat{w}_{j,t}$ of real user cost gap across different inputs j.

($\widetilde{w}_t > 0$), which AI may mitigate by improving supply-chain integration and procurement efficiency.

At the same time, AI can alter the equilibrium response of factor prices by changing task composition, input substitutability, and the effective elasticity of factor supply. This particularly relevant for specialized skills, robotics, and digital inputs (software, data, organizational capital). If AI raises returns to scale though automatization or allows firms to expand output using slack intangible capital, marginal costs may rise more slowly during expansions. Conversely, if AI adoption increases reliance on scarce complementary inputs—such as specialized chips, computing infrastructure, proprietary data, or high-skill labor—marginal costs may become more steeply increasing as economic activity accelerates.

Crucially, these margins need not move in the same direction, implying that AI can generate offsetting effects on $\widehat{w}_t$. For example, wages may become more flexible due to improved contracting technologies while remaining highly procyclical because labor supply of scarce complementary inputs—specialized skills, proprietary data, or computing capacity—is inelastic; conversely, wages may remain sluggish even as substitution possibilities expand. Based on the evidence available so far, the net effect of AI on the cyclical behavior of real input costs remains ambiguous.

Factor markets' frictions—Distortions in input markets that drive a gap between the real user cost of an input and the value of its marginal contribution to production also contribute to the cyclical movement of real marginal costs. In the equation above, $\widehat{\tau}_t := \tau_t - \tau_t^\star$ measures how frictional wedges, $\tau_t := w_t - mp_t$, are over and above those present in the flexible-price allocation $\tau_t^\star$.[4] The cyclicality of these wedges can be driven, among others, by changes in labor bargaining frictions, financing premia, adjustment costs, or markups embedded in user costs of intermediate inputs.

AI can generate cyclical variation in $\widehat{\tau}_t$ by altering the nature, intensity, and state-dependence of frictions in input markets. These effects need not be monotonic and may operate in opposite directions across markets and phases of the cycle. In labor

[4]Again, up to a first-order approximation around the natural benchmark, τ_t is a cost-share-weighted average $\sum_{j=1}^{J} c_j^\star \tau_{j,t}$ of wedges in different input markets. Note that $\tau_{j,t}$ capture average (or aggregate) wedges for the production factor j in the spirit of Chari et al. (2007). Aggregate wedges are different from misallocation wedge, which are arise due to cross-sectional heterogeneity in the frictions that individual firms face accessing factor markets (Hsieh and Klenow 2009). We capture these frictions inside the effeciency wedge $\widehat{a}_t$ discussed below.

markets, AI-enabled monitoring, performance evaluation, and algorithmic management may reduce informational frictions and weaken workers' bargaining power, compressing wage premia and reducing cyclical wage wedges. At the same time, AI can lower the cost of job search and application—through automated resume submission, matching platforms, and remote work—which may dramatically increase the volume of applicants per vacancy. This can raise screening and selection costs for firms, effectively increasing hiring frictions during expansions and generating positive cyclical wedges between wages and marginal products. Similarly, increased reliance on project-based or platform work may fragment employment relationships, making effective labor services more costly to scale despite flexible posted wages.

In capital markets, AI-driven improvements in credit scoring, risk assessment, and contract enforcement may reduce external finance premia and dampen the cyclicality of user costs. While this is a plausible baseline, AI may also amplify financial cyclicality by increasing the procyclicality of credit supply. In particular, algorithmic lending and model monocultures may lead intermediaries to over-relax credit standards during expansions and tighten them abruptly during downturns, often in a highly correlated manner across institutions. To the extent that AI-enabled finance raises leverage, accelerates credit booms, or synchronizes lending decisions, financing constraints can become more severe in busts despite higher underlying productivity. In such environments, the cyclical financing wedge rises and $\widehat{\tau}_t$ may increase even as measured productivity improves. We return to these mechanisms in Section 4.

Finally, AI can also affect wedges in intermediate-input markets by reshaping contracting technologies and supply-chain organization. Improved forecasting and inventory management may reduce delivery delays, renegotiation costs, and quantity-adjustment frictions, lowering cyclical wedges. At the same time, increased reliance on complex digital supply chains, proprietary platforms, or concentrated upstream providers may amplify contractual rigidities and raise effective input costs during periods of high demand.

Efficiency wedges—Finally, cyclical movements in production efficiency can either dampen cost pressures—when they reflect temporary productivity gains—or amplify them—when they entail transitory efficiency losses. In Equation (2), the term $\widehat{a}_t \equiv a_t - a_t^\star$ captures a *cyclical efficiency wedge*: the deviation of effective production efficiency from its flexible-price benchmark, holding fixed the underlying technological frontier and the real frictions that characterize the natural allocation.[5]

This wedge reflects forces that cause realized productivity to differ from its flexible-price counterpart even when the technological frontier is unchanged. A key mechanism is resource misallocation: nominal rigidities, sectoral bottlenecks, and adjustment frictions can prevent labor, capital, and intermediate inputs from flowing to their most productive uses, lowering effective productivity relative to the flexible-price allocation. Price and wage dispersion, congestion in factor markets, and imperfect reallocation across tasks or firms can all generate such efficiency losses, especially during periods of rapid structural change.

From the perspective of inflation dynamics, a negative efficiency wedge ($\widehat{a}_t < 0$) raises real marginal costs for a given level of activity, amplifying inflationary pressure even in the absence of strong demand or input-price growth. Conversely, improvements in allocative efficiency can partially offset cost pressures by allowing output to expand with smaller increases in marginal cost. Importantly, because $\widehat{a}_t$ reflects endogenous, state-dependent distortions rather than exogenous technology, its response to shocks—such as the diffusion of AI—need not be monotonic and can vary over the business cycle.

A common narrative surrounding AI diffusion emphasizes its cost-saving potential and its role in raising productivity. This narrative is often taken to imply, implicitly or explicitly, that AI will act as a deflationary force. While AI is widely regarded as a general-purpose technology that expands the economy's technological frontier—and therefore raises $a_t^{\star}$ over the long run—the implications for short-run inflation dynamics are far less straightforward. As explained, AI is disinflationary at cyclical horizons only insofar as it raises effective productivity relative to its flexible-price benchmark by more than it increases real unit input costs ($\widehat{mc}_t = \widehat{q}_t - \widehat{a}_t$). Productivity gains alone are therefore not sufficient to ensure disinflationary outcomes.

Crucially, it should not be taken for granted that AI diffusion generates a positive $\widehat{a}_t$ in the short run. The impact of AI on realized efficiency depends on adoption frictions, organizational adjustment, and complementarities with existing inputs, all of which are

[5] Effective productivity is defined as $a_t := (1 + \chi)\,\mathrm{tfp}_t + \tilde{a}_t$, where tfp_t denotes Hicks-neutral total factor productivity and χ captures amplification through returns to scale and utilization margins. The derivations reported in the Appendix.

subject to real and nominal rigidities. When prices and wages adjust sluggishly, firms face distorted relative prices that lead to inefficient allocation of inputs across producers and tasks, even as the underlying (potential) technological frontier continues to expand. These distortions raise the amount of inputs required to produce a given level of output, lowering realized efficiency relative to the flexible-price allocation. Thus, AI adoption can initially reduce (rather than increase) effective productivity, generating a negative $\widehat{a}_t$, despite ongoing improvements in the technological frontier. This non-monotonic pattern is consistent with the "productivity J-curve" emphasized by Brynjolfsson et al. (2021), whereby major technological innovations initially depress measured productivity before delivering sustained gains once complementary investments and organizational adjustments are completed.[6] These mechanisms become quantitatively more important during periods of rapid technological reorganization, such as the diffusion of AI.

Effects of AI on the passthrough of cyclical real marginal cost movements

A second channel though which the diffusion of AI can affect inflation dynamics is by affecting how marginal-cost fluctuations are transmitted into inflation through the slope of the Phillips curve, λ. The slope of the Phillips curve is determined by the interaction between nominal rigidities and real rigidities that shape firms' price-setting decisions. Following Gagliardone et al. (2025b) we summarize these forces through the decomposition:[7]

$$\lambda = \frac{(1-\theta)}{\theta}\,(1-\beta\theta)\,(1-\Omega)\,\Theta. \tag{3}$$

which highlights three distinct margins governing cost pass-through: the frequency of price adjustment (θ), discounting (β), strategic complementarities in price setting (Ω), and macroeconomic complementarities in production (Θ).

Nominal rigidities—Nominal rigidities limit the speed and extent to which firms can

[6]The authors emphasize how during early phases of diffusion, firms may incur substantial reorganization costs—such as learning, integration, data cleaning, workflow redesign, and experimentation—that temporarily divert resources away from production. Coordination failures arise as tasks are reallocated, and legacy systems coexist imperfectly with new technologies. In this phase, realized efficiency may fall short of its flexible-price benchmark even as trend TFP continues to improve. See also Bresnahan and Trajtenberg (1995) for a framework with long diffusion lags and complementarities in the diffusion and adoption of general purpose technologies.

[7]See Gagliardone et al. (2025b) for a derivation of the NKPC slope in an environment with oligopolistically competitive firms.

adjust prices in response to cost shocks, thereby governing how quickly marginal-cost fluctuations translate into inflation. To fix ideas, consider a standard Calvo (1983) pricing environment in which, in any given period, only a fraction $1-\theta$ of firms can reoptimize their prices, while the remaining firms keep prices unchanged. Within the Phillips-curve slope in Equation (3), the term $\frac{1-\theta}{\theta}$ captures the frequency of price adjustment: as price reoptimization becomes more frequent (lower θ), a larger share of firms can respond to cost shocks, increasing cost pass-through into inflation.[8]

Moreover, due to price stickiness, firms that do adjust internalize the fact that their prices are expected to remain in place for several periods. As a result, optimal reset prices are forward looking: firms choose prices to balance current marginal-cost conditions against expected future costs over the duration of price stickiness. This forward-looking pricing behavior implies that firms respond less aggressively to contemporaneous cost shocks when prices are expected to remain fixed for longer. The term $(1-\beta\theta)$ reflects the role of discounting in price-setting decisions. When firms place greater weight on future profits (higher β) and expect prices to remain fixed for longer (higher θ), current pricing decisions become more foreward looking—that is more sensitive to expected future marginal costs—reducing the strength of inflation responses to current movements in costs.

The diffusion of AI can affect nominal rigidities by reducing the managerial, informational, and computational costs of repricing. Automation, real-time demand forecasting, and algorithmic pricing tools can increase both the frequency and the state contingency of price adjustment. This corresponds to a decline in effective price stickiness (a lower θ), steepening the Phillips curve and allowing marginal-cost fluctuations to pass through more rapidly into inflation. While AI does not alter firms' fundamental time preferences, it can reshape the information set available to firms and the way expectations are formed, thereby affecting the effective role of β in price-setting decisions. In forward-looking pricing models, β governs the weight firms place on expected future marginal costs and inflation relative to current conditions. Improvements in data processing, forecasting accuracy, and real-time monitoring can increase firms' perceived

[8] Gagliardone et al. (2025a) develop a state-dependent pricing framework in which the frequency of price adjustment is endogenous and increasing in the magnitude of cyclical shocks (See also Nakamura and Steinsson 2010; Alvarez et al. 2022). The authors show that, in the absence of large aggregate disturbances, the Calvo pricing assumption provides a close approximation to firms' optimal pricing behavior (Gertler and Leahy 2008 Auclert et al. 2022).

ability to anticipate future economic conditions, effectively raising the importance of expected future outcomes in pricing decisions. In this sense, AI can increase the operational relevance of β, making inflation dynamics more expectation-driven and less tightly linked to contemporaneous economic slack.

At the same time, AI adoption may increase uncertainty about the medium run by accelerating structural change in market structure, technology, and competitive conditions. If firms become less confident about the persistence of future cost or demand conditions, they may discount the future more heavily in practice, reducing the influence of expected future marginal costs on current pricing decisions. In such environments, expectations may become less firmly anchored, and inflation may respond more sharply to current shocks, increasing volatility for given fundamentals. Thus, even holding deep preferences fixed, AI can affect inflation dynamics by altering how firms perceive, forecast, and discount future economic conditions.

Strategic price complementarities—Strategic complementarities arise when firms' desired prices depend on competitors' prices, leading individual firms to adjust prices less aggressively in response to changes in their current or anticipated production costs. The parameter $\Omega \in (0, 1)$ captures the strength of this channel: higher values of Ω imply that optimal reset prices place greater weight on competitors' prices and less on firms' own marginal costs. As a result, marginal-cost fluctuations translate less into relative price adjustments and, in the aggregate, into inflation.[9]

There is growing concern that AI may reshape competitive interactions in ways that strengthen strategic complementarities in pricing. First, AI adoption appears to be skewed toward larger firms, reflecting the importance of complementary intangible investments—such as data, organizational capital, and specialized skills—that are easier to finance and scale in large organizations (Calvino and Fontanelli 2023; OECD/BCG/INSEAD 2025; Lenzu et al. 2026). By reinforcing scale advantages, AI may increase market concentration and tilt pricing power toward incumbent firms, particularly

[9]In models with variable markups, Ω is increasing in the elasticity of desired markups with respect to relative prices. When demand elasticity is constant—as in the benchmark monopolistic competition model with Dixit–Stiglitz demand–markups are fixed, $\Omega = 0$, and strategic complementarities are absent. In that case, firms that reoptimize prices condition only on the discounted stream of their own marginal costs, competitors' prices are irrelevant, and the Phillips curve is steeper. When markups vary endogenously, pricing decisions become strategic complements and cost pass-through is attenuated. See Amiti et al. (2019) and Gagliardone et al. (2025b) for further discussion.

those that control compute capacity, data infrastructures, or cloud services. Second, AI improves information acquisition and processing, potentially increasing the extent to which firms condition their pricing decisions on competitors' prices. A growing literature on algorithmic pricing shows that learning algorithms can sustain supracompetitive pricing outcomes even in the absence of explicit communication (Calvano et al. 2020), a mechanism that naturally maps into stronger strategic complementarities.

At the same time, these effects may be counterbalanced by opposing competitive forces. Widespread AI adoption can increase price transparency, reduce search and switching costs, and lower barriers to entry—particularly in platform-based and digital markets—thereby compressing markups and weakening incumbents' pricing power. In such environments, strategic complementarities may attenuate and increasing cost pass-through into inflation. Whether AI ultimately flattens or steepens the Phillips curve through this channel therefore depends on which of these opposing forces dominates, an outcome that is likely to vary across sectors, market structures, and stages of AI adoption.

Macroeconomic complementarities—Even when individual firms face identical price-setting frictions, the aggregate pass-through of cost shocks into inflation can be dampened by general-equilibrium feedbacks. These forces are captured by the coefficient $\Theta \leq 1$, which summarizes the role of macroeconomic complementarities in the transmission of marginal-cost fluctuations into prices. Macroeconomic complementarities arise from features of the production environment—such as decreasing returns to scale, shared factor markets, and input-output linkages—that cause firms' costs to move together in response to aggregate disturbances, once all general-equilibrium feedbacks are taken into account.

Intuitively, when aggregate output expands, higher factor prices, tighter input markets, and rising intermediate-input costs affect all firms simultaneously. Because marginal costs rise broadly rather than idiosyncratically, relative cost differences across firms are compressed. As a result, the gap between a firm's optimal price and the aggregate price level increases by less, leading to smaller desired relative price adjustments.[10]

[10] The key point is that firms adjust prices to restore *relative* markups, not to offset changes in the level of marginal costs per se. When a cost increase is largely idiosyncratic, raising prices improves a firm's relative position. By contrast, when cost shocks are aggregate and shared across firms, relative marginal costs—and hence relative markups—move little. Expected inflation and general-equilibrium feedbacks then partially restore markups over time, reducing the marginal benefit of aggressive immediate price adjustment. Macroeconomic complementarities therefore dampen inflation not because marginal costs rise

These general-equilibrium feedbacks—captured by a parameter $\Theta < 1$—dampen cost pass-through and flatten the Phillips curve, implying that inflation responds less to aggregate activity than under partial-equilibrium pricing.

AI can affect inflation dynamics through this channel by reshaping the structure of production and the nature of general-equilibrium feedbacks. Beyond raising productivity, AI may alter effective returns to scale, task modularity, and network dependencies in supply chains and digital infrastructure. To the extent that AI increases effective returns to scale—by enabling firms to expand output using shared platforms, software, or data—macroeconomic complementarities weaken and Θ rises toward one, increasing the sensitivity of inflation to real activity. Conversely, if AI adoption intensifies input-output linkages or increases reliance on common upstream inputs and infrastructures, aggregate cost movements may become more synchronized, strengthening macroeconomic complementarities and dampening cost pass-through into inflation.

Taken together, these considerations suggest that AI reshapes—rather than eliminates—the inflation–activity tradeoff, and it does so through multiple margins that need not move in the same direction. Real-side effects operating through marginal costs, pricing frictions governing pass-through, and expectation-formation channels may offset or reinforce one another, rendering the net impact of AI on the slope of the Phillips curve ambiguous *ex ante*. Even if individual mechanisms are well understood in isolation, their joint effect depends on how AI alters the interaction between production technologies, pricing behavior, and market structure.

Importantly, this ambiguity is likely to be amplified by heterogeneity. While the analysis above abstracts from cross-industry variation and treats the "aggregate Phillips curve" as a representative aggregate relationship, in practice inflation dynamics reflect the aggregation of heterogeneous industry-level Phillips curves. AI may not only change the structural parameters governing price adjustment and cost pass-through at the firm or industry level, but also imply a quick and significant reallocation economic activity

less, but because cost increases are common across firms, weakening incentives to adjust prices aggressively. In the appendix, we derive Θ as a function of the elasticity of substitution across producers, the sensitivity of marginal costs to output, and the strength of strategic complementarities (see also Gagliardone et al. 2025b). Note that this mechanism is distinct from the role of χ, which governs how marginal cost responds to changes in the level of economic activity.

across sectors with systematically different pricing frictions, production elasticities, and exposure to cost shocks. As a result, the slope of the aggregate Phillips curve may evolve more rapidly than implied by within-industry adjustments alone, reflecting both parameter changes and compositional shifts in economic activity.[11]

2.2 Cyclical transmission of aggregate demand disturbances and expectations

In a canonical business-cycle model, the aggregate demand side is captured by the dynamic investment-saving (IS) relation, a log-linearized Euler equation that relates current and expected aggregate expenditure, expressed as deviations from natural levels, to the real interest rate gap:

$$y_t = \mathbb{E}_t\{y_{t+1}\} - \frac{1}{\sigma}\left(r_t - r_t^\star\right) + \varepsilon_t^{IS}, \tag{4}$$

where the ex-ante real rate is given by $r_t := i_t - \mathbb{E}_t\{\pi_{t+1}\}$ is the real interest rate and $r_t^\star$ is the natural (Wicksellian) real rate consistent with zero output gap; σ governs the sensitivity of demand to real interest rate gaps; ε_t^{IS} collects demand disturbances.

Mirroring the approach in the previous section, for the purposes of analyzing the short-run inflationary effects of AI, we abstract from demand-side forces that shift the natural rate $r_t^\star$ or alter preference parameters such as σ, which are equilibrium objects discussed in Section 3. Instead, we will focus on a distinct and empirically relevant channel through which AI affects aggregate demand at cyclical horizons: expectations.

The diffusion of AI can alter agents' beliefs about future productivity, income, and profitability even before these changes materialize in measured costs or potential output.[12] Anticipated AI-driven gains raise expected future income and investment

[11] In an environment with cross-industry heterogeneity in nominal rigidities and strategic complementarities, aggregation implies that the cost-based NKPC can be written as

$$\pi_t = \lambda\,\widehat{mc}_t + \mathrm{Cov}\big(\lambda_i, \widehat{mc}_{i,t}\big) + \beta\,\mathbb{E}_t\{\pi_{t+1}\},$$

where $\lambda := \int \lambda_i\,di$ is the average slope and $\lambda_i := \frac{(1-\theta_i)(1-\beta\theta_i)}{\theta_i}(1-\Omega_i)\Theta_i$. Aggregate inflation therefore depends not only on the average pass-through but also on cross-sectional covariances between marginal-cost fluctuations and industry-specific pricing frictions, such as $\mathrm{Cov}(\widehat{mc}_{i,t}, \theta_i)$, $\mathrm{Cov}(\widehat{mc}_{i,t}, \Omega_i)$, and $\mathrm{Cov}(\widehat{mc}_{i,t}, \Theta_i)$. AI-driven reallocation toward sectors with more flexible pricing, stronger complementarities, or tighter cost pressures can thus change aggregate inflation dynamics even if within-industry parameters remain unchanged. These aggregation effects may be further amplified in the presence of input–output linkages.

[12] Formally, anticipated AI adoption can be represented as a news shock to future demand conditions,

returns, stimulating current consumption and investment through intertemporal substitution. As a result, aggregate demand may expand and the output gap may open even if contemporaneous productivity and potential output remain unchanged. This expectations-driven expansion increases utilization and tightens input markets, generating upward pressure on inflation through the Phillips curve.

These considerations raise a natural question: should AI-driven demand news shocks be expected to differ systematically from traditional demand shocks? In reduced form, both operate by shifting the output gap for a given real interest rate gap and can therefore appear similar on impact. The distinction becomes economically relevant once one recognizes that AI-related demand pressure reflects revisions in beliefs about future productivity rather than an exogenous contemporaneous spending impulse. This difference affects the timing and comovement of output, costs, and inflation, complicates real-time identification, and increases the sensitivity of inflation dynamics to how expectations are formed and updated.

A first implication concerns timing and comovement. Because news shocks work through expectations about future conditions, they tend to generate more front-loaded movements in current demand than standard "spot" demand shocks. Anticipated AI-driven gains stimulate consumption and investment today through the entire expected path of future income and returns, even if productivity improvements have not yet materialized. As a result, output and marginal costs may rise before supply-side efficiency gains are reflected in measured productivity. Whether the resulting inflationary pressure proves temporary or persistent depends on how quickly productivity gains are realized relative to this initial demand expansion.

This logic connects directly to the efficiency-wedge discussion above. Even when AI is expected to lower production costs in the long run, its diffusion can generate short-run inflationary pressure if expectations-driven demand responds more rapidly than effective productivity. Short-run inflation outcomes therefore depend on the relative timing and strength of two opposing forces: front-loaded demand driven by expectations and the gradual realization of cost-reducing productivity gains on the supply side. When the former dominates, inflation can rise temporarily despite anticipated technological

captured either by revisions in expectations of future output gaps $\mathbb{E}_t\{y_{t+1}\}$ or by innovations to the expected future path of IS disturbances ε^{IS}_{t+j} for $j \geq 1$, rather than by a contemporaneous demand shock ε^{IS}_t.

progress.[13]

A second, policy relevant implication concerns "identification". AI-driven demand news is fundamentally about future supply, but it initially manifests as demand pressure. In real time, this makes it difficult to distinguish from conventional demand overheating, particularly when measured productivity responds sluggishly due to adoption frictions, reorganization costs, or misallocation. As discussed in the context of efficiency wedges, realized productivity may temporarily lag behind its flexible-price benchmark even as expectations improve. In such environments, inflationary pressure may reflect intertemporal substitution ahead of future capacity rather than excess demand relative to long-run productive potential, posing a challenge for policy calibration.

3 Effects of AI on long-run structural transition

In the log run, a diffused adoption of AI technologies are best understood not as changes in cyclical transmission, but as shifts in the benchmarks around which monetary policy stabilizes the economy and as changes in intertemporal allocation.

3.1 Potential output and the natural rate of interest

The long-run implications of artificial intelligence operate through its effects on the economy's natural benchmarks—the natural (flexible-price) level of output $y_t^\star$ and the natural real interest rate $r_t^\star$. The natural level of output $y_t^\star$ is the level of economic activity that would prevail in the absence of nominal rigidities; it reflects underlying fundamentals such as productivity, labor supply, production technologies, and market structure. The natural real rate of interest $r_t^\star$ is the real interest rate consistent with that same flexible-price allocation. Equivalently, it is the real rate that supports a zero output gap and stable inflation in the sticky-price economy. As nominal rigidities dissipate over time, the economy converges toward this flexible-price equilibrium, in which inflation is stable and real activity is determined solely by preferences, technology, and market structure.

[13] The distinction between anticipated and unanticipated AI adoption is emphasized in a general-equilibrium setting by Aldasoro et al. (2024a), who show that anticipated AI adoption tends to generate a temporary inflationary response through demand, whereas unanticipated productivity shocks are more likely to be disinflationary on impact.

In standard business-cycle frameworks, the relationship between $y_t^\star$ and $r_t^\star$ is embedded in the household Euler equation and, by extension, in the dynamic IS relation (Equation 4). Given a path for flexible-price output and consumption, the natural real rate adjusts to equate desired saving and investment over time. Higher expected growth in potential output raises expected consumption growth and the marginal return to investment, thereby increasing the real interest rate required to induce households to postpone expenditure. Conversely, forces that depress investment demand or increase precautionary saving put downward pressure on the natural rate. Movements in $y_t^\star$ therefore map naturally into movements in $r_t^\star$ through intertemporal allocation decisions.[14]

Monetary policy stabilizes the economy around these evolving benchmarks rather than attempting to influence them directly. Abstracting from unconventional interventions, monetary policy affects inflation and real activity by influencing deviations of the policy real rate from $r_t^\star$, not through the level of $r_t^\star$ itself—a mechanism made explicit by the IS equation in (4).

Policy errors therefore arise not because the natural rate moves, but because the central bank fails to update its assessment of those movements, creating persistent gaps between the policy real rate and the evolving neutral rate. These considerations map directly into standard policy rules. For example, under a conventional Taylor-style rule,

$$i_t = r_t^\star + \bar{\pi} + \phi_\pi(\pi_t - \pi^\star) + \phi_x x_t, \tag{5}$$

changes in $r_t^\star$ require corresponding adjustments to the intercept of the policy rule. If the natural rate rises and policy does not follow, policy becomes unintentionally expansionary; if the natural rate falls and policy does not follow, policy becomes unintentionally contractionary. Reacting to inflation and activity alone is therefore insufficient when benchmark estimates are persistently mismeasured.

AI matters for monetary policy because it can affect both the level and the growth rate of potential output, as well as the intertemporal trade-offs that determine $r_t^\star$, through multiple, conceptually standard channels. By raising trend productivity and expanding

[14] In the canonical New Keynesian model, the natural real rate of interest is pinned down by the Euler equation evaluated at the flexible-price allocation. Abstracting from preference shocks and wedges, this implies an approximate log-linear relationship of the form $r_t^\star = -\log\beta + \sigma\,\mathbb{E}_t\{\Delta y_{t+1}^\star\}$, where β is the subjective discount factor and $\mathbb{E}_t\{\Delta y_{t+1}^\star\}$ denotes expected growth in natural output. Higher expected growth raises the natural real rate by increasing the real return required to induce households to postpone expenditure. See Woodford (2011) and Galí (2015) for formal derivations.

profitable investment opportunities, AI can increase expected consumption growth and the marginal return to capital, putting upward pressure on both potential output and the natural real rate. At the same time, AI may reshape market structure, increase concentration, or alter the distribution of rents in ways that dampen aggregate investment demand. AI-driven uncertainty, labor displacement risk, or distributional effects may also raise precautionary saving, exerting downward pressure on the natural rate. The effectiveness of monetary policy in an AI-intensive economy therefore hinges on correctly tracking changes in natural benchmarks and adjusting the stance of policy accordingly, rather than reacting mechanically to AI-driven changes in activity or prices.

Importantly, while theory provides clear guidance on how $r_t^\star$ should enter policy decisions (Woodford 2011; Galí 2015; Clarida et al. 1999), both $y_t^\star$ and $r_t^\star$ are unobserved and subject to substantial real-time uncertainty (Laubach and Williams 2003; Holston et al. 2017). Rapid structural change driven by AI may exacerbate this challenge by making it harder to distinguish cyclical movements from shifts in underlying trends. This increases the value of robustness, gradualism, and careful inference in policy design, even as the conceptual role of natural benchmarks remains unchanged.

3.2 Intertemporal elasticity and aggregate demand sensitivity

A second long-run channel through which AI affects the conduct of monetary policy operates through the sensitivity of aggregate demand—consumption and investment—to movements in real interest rates.

This channel is captured by the parameter σ, the intertemporal elasticity of substitution in consumption, in the IS equation (4). In textbook New Keynesian models, σ is a primitive object deriving from household preferences, capturing the curvature of utility and the willingness to substitute consumption across time.[15] More broadly, however, σ should be interpreted as a reduced-form equilibrium object summarizing all frictions that limit intertemporal reallocation. These include borrowing constraints, income risk, consumption-smoothing mechanisms, and information frictions.

As we will discuss in Section 4, AI may affect σ through several slow-moving

[15]In textbook NK models, the intertemporal elasticity of substitution σ enters household preferences together with the inverse Frisch elasticity φ: $U(C_t, N_t) = \frac{C_t^{1-\sigma}-1}{1-\sigma} - \frac{N_t^{1+\varphi}}{1+\varphi}$. Thus, σ governs the curvature of utility and households' willingness to substitute consumption intertemporally.

structural channels. Improvements in financial intermediation—such as enhanced credit scoring and risk assessment—can relax borrowing constraints and improve consumption smoothing, increasing the responsiveness of demand to real interest rates. AI may also reduce income risk through better forecasting and risk management, lowering precautionary saving and raising effective intertemporal substitution. In addition, AI can reduce investment planning and implementation costs, making investment demand more elastic with respect to expected returns. Finally, improvements in forecasting and financial advising may lengthen planning horizons and increase confidence about future outcomes, further amplifying intertemporal substitution.

Note that, unlike structural transformations that operate through shifts in $y^\star$ and $r^\star$, these effects of AI diffusion operating through σ matter for central banks because they shape the strength of monetary policy transmission to real activity, rather than the equilibrium long-run level of economic activity itself.

4 Implications of AI diffusion for financial stability

While price stability remains the primary objective of central banks, in most jurisdictions they also play a critical role in safeguarding financial stability, either by mandate or by necessity. The rapid diffusion of AI technologies—particularly generative AI—will inevitably intersect with this objective by reshaping financial intermediation, risk assessment, and the origination and propagation of asset-price shocks. Although these effects lie outside the frameworks discussed in the previous sections—cyclical transmission and long-run changes in fundamentals—they interact with them by altering the financial landscape through which monetary policy is transmitted to the real economy.

4.1 Impact on financial intermediaries

A growing body of evidence shows that AI is increasingly embedded across core segments of the financial system that are central to financial intermediation and, *a fortiori*, financial stability: lending, insurance, and asset management.

In lending, the integration of GenAI with existing machine-learning-based credit scoring and underwriting schemes has the potential to expand access to credit and improve risk assessment, particularly for borrowers with limited credit histories (Berg et al. 2020; Fuster et al. 2022). However, these benefits should be weighed against the nontrivial risks posed by these same tools. First, widespread adoption of these technologies by financial intermediaries may amplify the procyclicality of credit cycles by increasing the sensitivity of credit supply to real-time signals or market sentiment. Moreover, opaque and complex models trained on historical data may embed biases or perform poorly under structural change, leading to correlated mispricing of risk across institutions and increasing systemic vulnerability (Aldasoro et al. 2024b; International Monetary Fund 2024).

In insurance, AI enhances risk classification, pricing, and claims management, potentially improving efficiency and reducing fraud. However, finer risk segmentation may weaken traditional risk pooling and increase exposure to tail risks, particularly when extreme events fall outside the training data used by models (Balasubramanian et al. 2018; European Insurance and Occupational Pensions Authority 2021. These developments raise important questions about the long-run insurability of certain risks and the resilience of insurers' balance sheets under stress (International Monetary Fund, 2024).

In asset management, AI is increasingly used for portfolio optimization, algorithmic trading, sentiment analysis, and risk management (Gensler, 2023). While these tools can improve information processing and execution speed, they may also contribute to faster and more synchronized portfolio adjustments, increasing market volatility and the risk of abrupt price movements (Kirilenko et al. 2017).

4.2 The topology of the financial system and the transmission of monetary policy

The adoption of AI by the financial sector may alter not only individual institutions' behavior but also the very topology of the financial system, increasing interconnectedness, the speed of adjustment, and shaping competitive dynamics. These changes have direct implications for the transmission and effectiveness of monetary policy.

A large literature emphasizes that monetary policy affects the real economy not only—or perhaps not even primarily—through the risk-free real interest rate, but through financial intermediaries whose balance sheets, funding constraints, and risk-bearing capacity shape credit spreads and the supply of credit to households and firms (Bernanke and Gertler, 1989; Bernanke et al., 1999; Gertler and Karadi, 2011). More recently, this perspective has been extended to highlight the growing role of market-based finance—including securitization, private credit, and CLOs—in transmitting monetary policy through asset prices, leverage, and risk premia (Adrian and Shin 2010; Di Maggio et al. 2020).

A simple way to connect financial conditions back to the canonical New Keynesian demand block is to interpret the relevant real rate in the IS equation as the *effective real rate* faced by households and firms. In particular, the IS relation can be written as

$$y_t = \mathbb{E}_t\{y_{t+1}\} - \frac{1}{\sigma}\Big(\underbrace{r_t + \tau_t^{IS}}_{\text{effective real rate}} - r_t^{\star}\Big) + \varepsilon_t^{IS}, \tag{6}$$

where τ_t^{IS} denotes a a financial wedge—such as a credit spread or external finance premium—that reflects intermediary balance-sheet conditions, funding constraints, and market liquidity.[16] This representation clarifies why the topology of the financial system matters for monetary policy: even if the central bank influences market rates through conventional or unconventional operations, movements in s_t can materially alter the effective stance of policy as perceived by the private sector.

To fix ideas, consider a central bank that sets the nominal policy rate in response to inflation and economic slack, conditional on its assessment of $r_t^{\star}$, according to the Taylor rule in Equation (5). For a given setting of the policy rate i_t, different realizations of the spread τ_t^{IS} can therefore imply very different degrees of effective monetary accommodation or tightening, depending on how financial conditions respond to the policy action.

AI can affect monetary transmission through financial channels by altering the behavior of s_t itself. In particular, AI adoption may make spreads more volatile, more state-dependent, or more correlated across institutions, increasing the likelihood that financial conditions tighten abruptly even when inflation is near target and measured output gaps are small. These effects operate through two closely related mechanisms.

First, AI may reshape competitive dynamics and market structure within

[16] Just as $\widehat{\tau}_t$ captures wedges between factor prices and marginal products, τ_t^{IS} captures, in reduced form, wedges between the policy rate and the intertemporal price relevant for private spending decisions. In the spirit of Bernanke et al. (1999) and Gertler and Karadi (2011), the idea that monetary policy transmits to the real economy in part through endogenous spreads that depend on intermediary net worth, leverage constraints, and risk-bearing capacity.

intermediary sectors—for example, by strengthening economies of scale in data and model deployment, increasing reliance on a small number of technology providers, or accelerating the diffusion of similar risk signals—thereby affecting intermediaries market power, risk-taking incentives, and the elasticity of spreads and credit supply with respect to policy rates. Second, AI may be adopted differentially across intermediaries with heterogeneous funding structures and regulatory constraints. If credit intermediation shifts toward segments characterized by weaker capital regulation, greater maturity transformation, or distinct risk-management practices, the mapping from policy rates to aggregate financial conditions may change, altering both the strength and the timing of monetary policy transmission to real activity.[17]

4.3 Financial infrastructures and systemic interactions

Another central concern for financial stability is that widespread AI adoption may promote "model monocultures," in which many financial institutions rely on similar datasets, algorithms, or foundation models to make decisions. When technological penetration is high and AI service provision is concentrated, institution-level efficiencies can also be a source of system-wide vulnerabilities (Financial Stability Board 2017; Financial Stability Board 2024).

AI may amplify herding behavior and endogenous correlation. If many market participants react to the same signals or model outputs, asset prices may deviate persistently from fundamentals and then adjust sharply once beliefs shift (International Monetary Fund 2024). This mechanism resembles well-known amplification channels in financial crises, in which relatively small shocks are endogenously magnified through balance-sheet feedbacks and expectations, as formalized in the financial accelerator literature (Bernanke et al. 1999) and in models of belief-driven boom–bust cycles (Benigno and Fornaro 2018).

Other considerations apply to the integration of AI technologies in the payment system. AI-driven automation, fraud detection, and real-time monitoring can improve efficiency and operational resilience. At the same time, increased reliance on complex algorithms and a small number of technological providers may heighten operational and

[17] A growing literature documents a secular shift toward market-based finance and private credit, with important implications for both the strength and the timing of monetary policy transmission to real activity (Fleckenstein et al. 2025; Ivashina 2025).

cyber risks, creating potential single points of failure with systemic consequences in the event of outages, cyber incidents, or model failures (McMahon et al. 2024; Kazinnik and Brynjolfsson 2025).

Finally, the opacity of complex AI systems complicates risk management and supervision. Errors, biases, or breakdowns in widely used models can become common shocks rather than idiosyncratic disturbances, so that risks appearing diversified at the micro level are highly correlated at the system level (International Monetary Fund 2024; Basel Committee on Banking Supervision 2020; Gensler 2023). These features raise the likelihood that institution-level efficiencies translate into system-wide fragility, reinforcing the importance of financial stability considerations for central banks in an AI-intensive economy (Financial Stability Board, 2017; International Monetary Fund, 2024).

4.4 Asset valuations and stock-market crash risk

Finally, one should not understate the potential financial stability risks that the diffusion and adoption of AI may pose through their effects on asset valuations and expectations, particularly in equity markets. The rapid progress of generative AI technologies has generated substantial optimism about future productivity gains, profitability, and long-run economic growth. This optimism has been reflected in elevated equity valuations, an increasing concentration of market capitalization in firms perceived to be at the technological frontier of AI adoption, and very large investment flows aimed at expanding the capacity required to meet anticipated demand for AI-related services.

As discussed in Sections 2 and 3, to the extent that these valuations and investment flows reflect rational expectations, they can be understood as part of the transition to a higher-productivity long-run equilibrium. That transition may nonetheless entail temporary challenges for price stability along the adjustment path, as well as persistent risks related to market concentration and the allocation of rents. However, this transition is inherently uncertain—if not in its ultimate direction, then at least in the timing and speed with which it unfolds. AI-driven productivity gains may materialize more slowly than anticipated, may be unevenly distributed across sectors, or may be partially offset by increased market power accruing to the winners of the AI arms race.

These observations are particularly relevant for the current policy debate. If expectations about the pace of this transition are revised abruptly, asset prices may adjust sharply, giving rise to stock-market corrections or crashes with broad macro-financial consequences, even when inflation is close to target and output gaps are small (Bank for International Settlements 2025). Such consequences may include a significant tightening of financial conditions, impairment of financial intermediaries' balance sheets, and disruptions to credit provision.

Excessive leverage and risk-taking behavior may further amplify downside risks (International Monetary Fund 2024; Gopinath 2025; The Economist 2025), particularly when the financing that feeds asset valuations and investment is intermediated through nonbank financial institutions. Recent research emphasizes the rapid expansion of private credit and other nonbank financing channels—often characterized by greater opacity, leverage, and maturity transformation—with banks frequently acting as liquidity or credit backstops (International Monetary Fund 2024; Acharya et al. Forthcominga; Acharya et al. Forthcomingb).

Owing to the nature of their business models, such firms producers and adopters of AI technologies tend to rely less on traditional bank lending and more on private credit and other forms of nonbank financing.[18] Although more data are needed before drawing firm conclusions, this raises a salient concern that the recent AI-led investment wave may have generated substantial private credit exposures for both banks and nonbank financial institutions to AI-related firms, including developers, adopters, and firms supplying complementary inputs and infrastructure.

Similar considerations apply to potential sources of fragility embedded in the balance sheets of insurance companies. Over the past decade, insurers have increasingly shifted their asset portfolios toward privately placed debt and securitized corporate loans (CLOs) issued by nonfinancial firms (Fringuellotti and Santos 2025; Fournier et al. 2024). These instruments are typically less liquid, more opaque, and subject to weaker market discipline than publicly traded bonds. While systematic evidence on the sectoral composition of insurers' private credit exposures remains limited, it is plausible that a nontrivial share of this exposure is to firms at the frontier of AI production and adoption, whose characteristics and funding needs align naturally with the long

[18] These observations are consistent with evidence showing that firms with high intangible capital intensity make limited use of bank debt and rely disproportionately on equity and nonbank financing; see Falato et al. (2022) and Jang et al. (2025).

investment horizons and return profiles sought by insurance companies.

Taken together, these financial linkages interacting with leveraged financial structures whose contours are not yet fully understood increase the potential for valuation-driven shocks to propagate across the entire financial system.

5 Policy Implications and concluding remarks

Artificial intelligence represents a technological transformation whose macroeconomic implications extend well beyond any single transmission channel. For central banks, the relevant question is therefore not whether AI matters, but how it matters for the conduct of monetary policy—specifically, how AI reshapes inflation dynamics, equilibrium benchmarks, and financial conditions in ways that affect stabilization trade-offs. To organize these issues, this paper has distinguished between three interrelated dimensions of direct relevance for monetary authorities: cyclical transmission, structural transition, and financial stability.

The canonical New Keynesian monetary framework remains a useful and flexible organizing device for analyzing these effects. Viewed through this lens, the diffusion of AI does not call for a redefinition of monetary policy objectives, nor does it imply that central banks should respond mechanically to technological innovation or asset prices per se. Instead, by altering the mapping from economic conditions into inflation and financial risks, the diffusion AI will likely complicate the interpretation of familiar indicators and the application of standard policy rules, requiring greater caution and judgment in real-time policy assessment.

A central policy implication of the supply-side analysis of shock transmission developed in the paper is that AI can alter inflation dynamics even when traditional measures of economic slack appear unchanged. By reshaping the elasticity of marginal costs with respect to output, the cyclical behavior of real input prices, and the degree of cost pass-through into prices, AI may weaken—or, in some sectors, strengthen—the link between inflation and standard indicators such as the output gap or labor market tightness. For policymakers, this implies that inflation may become a less reliable real-time signal of cyclical conditions, and that a given movement in activity may generate inflationary pressure that differs materially from historical experience.

More fundamentally, AI may alter the inflation–real activity trade-off itself. If AI allows output to expand with smaller increases in real marginal cost—by lowering the elasticity of production costs with respect to activity—the effective slope of the output-gap Phillips curve becomes flatter. In such an environment, a given movement in economic slack generates weaker inflationary pressure, making inflation less informative about contemporaneous demand conditions. For monetary policy, the implication is not a mechanical change in policy-rules, but a shift in the stabilization trade-off: safeguarding price stability may require smaller movements in economic slack, or tighter financial conditions, than under historical relationships. This increases the value of robust policy strategies that perform well under parameter uncertainty and heightens the importance of real-time inference about the forces driving movements in real production costs. In practice, this calls for greater emphasis on cost-side diagnostics—including real marginal cost proxies, input-market conditions, and pricing behavior—rather than exclusive reliance on reduced-form Phillips curve relationships.

A related challenge concerns real-time inference about equilibrium benchmarks, including potential output and the natural rate of interest. Rapid AI-driven structural change can induce persistent uncertainty about these objects, increasing the risk of policy miscalibration. As emphasized in the analysis, short-run productivity dynamics associated with AI adoption need not align with long-run technological gains: transitional frictions, reorganization costs, and misallocation can temporarily depress effective productivity even as the technological frontier expands. For monetary policy, this underscores the importance of distinguishing between cyclical inflationary pressures arising from demand and costs, and movements driven by shifting benchmarks that should not elicit a stabilization response. At the same time, AI may make monetary policy more powerful in affecting real economic activity. Policy actions may transmit more rapidly to inflation and activity as information flows accelerate, expectations adjust more quickly, and pricing decisions respond with shorter lags. In an environment where cost dynamics, pricing behavior, and financial conditions evolve jointly and nonlinearly, policy errors are less likely to be absorbed gradually and more likely to be amplified.

The diffusion of AI also raises new financial stability considerations that interact closely with monetary policy. By accelerating information processing, encouraging reliance on common technologies and models, and interacting with leveraged balance sheets—often outside the traditional banking system—AI may increase the likelihood that financial stress originates from expectations-driven valuation dynamics rather than from conventional macroeconomic imbalances. Sudden asset-price corrections can therefore tighten financial conditions rapidly, impair intermediaries' balance sheets, and disrupt credit provision, even in environments where inflation appears well contained. These dynamics strengthen the case for a policy framework that clearly distinguishes between tools aimed at price stability and those aimed at financial stability, while recognizing that the two domains may interact more tightly in an AI-intensive economy.

A useful way to synthesize the core insights of this paper is through a simple analogy. If the economy is a car, AI upgrades the engine by raising potential speed—through higher productivity, expanded capacity, improved information, and altered financial intermediation—while simultaneously making the steering more sensitive by reshaping inflation transmission, shifting policy benchmarks, and amplifying financial feedbacks. The task of central banks is not to slow or accelerate the engine, but to adjust the steering: calibrating policy in a way that maintains macroeconomic stability as the structure of the economy evolves. From this perspective, successful monetary policy hinges not on reacting to AI per se, but on maintaining clarity about what policy can and cannot control, improving real-time inference about costs and benchmarks, and designing robust strategies that perform well under heightened structural uncertainty. In this sense, AI reinforces—rather than overturns—a central lesson of modern monetary policy: effective stabilization requires a deep and continuously updated understanding of the economy to which policy is applied.

References

Daron Acemoglu and Pascual Restrepo. Artificial intelligence and jobs. *Journal of Economic Perspectives*, 34(3):30–50, 2020.

Viral V Acharya, Nicola Cetorelli, and Bruce Tuckman. Where do banks end and nbfis begin? *Review of Corporate Finance Studies*, Forthcominga.

Viral V Acharya, Nicola Cetorelli, and Bruce Tuckman. Transformed intermediation:

Credit risk to nbfis, liquidity risk to banks. *Journal of Finance Insights and Perspectivess*, Forthcomingb.

Tobias Adrian and Hyun Song Shin. Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437, 2010.

Philippe Aghion, Benjamin Jones, and Charles I. Jones. Artificial intelligence and economic growth. In Ajay Agrawal, Joshua Gans, and Avi Goldfarb, editors, *The Economics of Artificial Intelligence: An Agenda*, pages 237–282. University of Chicago Press, 2019.

Inaki Aldasoro, Sebastian Doerr, Leonardo Gambacorta, and Daniel Rees. The impact of artificial intelligence on output and inflation. BIS Working Paper 1179, Bank for International Settlements, 2024a.

Iñaki Aldasoro, Leonardo Gambacorta, Anton Korinek, Vatsala Shreeti, and Merlin Stein. Intelligent financial system: how ai is transforming finance. Technical report, Bank for International Settlements, 2024b.

Fernando Alvarez, Francesco Lippi, and Aleksei Oskolkov. The macroeconomics of sticky prices with generalized hazard functions. *The Quarterly Journal of Economics*, 137(2): 989–1038, 2022.

Mary Amiti, Oleg Itskhoki, and Jozef Konings. International shocks, variable markups, and domestic prices. *The Review of Economic Studies*, 86(6):2356–2402, 2019.

Andrew Atkeson and Ariel Burstein. Pricing-to-market, trade costs, and international relative prices. *American Economic Review*, 98(5):1998–2031, 2008.

Adrien Auclert, Rodolfo D Rigato, Matthew Rognlie, and Ludwig Straub. New pricing models, same old phillips curves? Technical report, National Bureau of Economic Research, 2022.

Ramnath Balasubramanian, Ari Libarikian, and Doug McElhaney. Insurance 2030—the impact of ai on the future of insurance. *McKinsey & Company*, pages 1–10, 2018.

Bank for International Settlements. Annual economic report 2025: Chapter 3. financial conditions in a changing global financial system. Technical report, Bank for International Settlements, June 2025.

David R Baqaee, Emmanuel Farhi, and Kunal Sangani. The supply-side effects of monetary policy. *Journal of Political Economy*, 132(4):1065–1112, 2024.

Basel Committee on Banking Supervision. Sound practices: Model risk management. Technical report, Bank for International Settlements, 2020.

Gianluca Benigno and Luca Fornaro. Stagnation traps. *Review of Economic Studies*, 85(3): 1425–1470, 2018.

Tobias Berg, Valentin Burg, Ana Gombović, and Manju Puri. On the rise of fintechs: Credit scoring using digital footprints. *The Review of Financial Studies*, 33(7):2845–2897, 2020.

Ben S. Bernanke and Mark Gertler. Agency costs, net worth, and business fluctuations. *American Economic Review*, 79(1):14–31, 1989.

Ben S. Bernanke, Mark Gertler, and Simon Gilchrist. The financial accelerator in a quantitative business cycle framework. *Handbook of Macroeconomics*, 1:1341–1393, 1999.

Timothy F Bresnahan and Manuel Trajtenberg. General purpose technologies 'engines of growth'? *Journal of econometrics*, 65(1):83–108, 1995.

Erik Brynjolfsson, Daniel Rock, and Chad Syverson. The productivity j-curve: How intangibles complement general purpose technologies. *American Economic Journal: Macroeconomics*, 13(1):333–372, 2021.

Erik Brynjolfsson, Danielle Li, and Lindsey Raymond. Generative ai at work. NBER Working Paper 31161, National Bureau of Economic Research, 2023.

Emilio Calvano, Giacomo Calzolari, Vincenzo Denicolò, and Sergio Pastorello. Artificial intelligence, algorithmic pricing, and collusion. *American Economic Review*, 110(10): 3267–3297, 2020.

Flavio Calvino and Luca Fontanelli. A portrait of ai adopters across countries. *Documents de travail de l'OCDE sur la science, la technologie et l'industrie*, 2023.

Guillermo A Calvo. Staggered prices in a utility-maximizing framework. *Journal of Monetary Economics*, 12(3):383–398, 1983.

Varadarajan V Chari, Patrick J Kehoe, and Ellen R McGrattan. Business cycle accounting. *Econometrica*, 75(3):781–836, 2007.

Richard Clarida, Jordi Gali, and Mark Gertler. The science of monetary policy: A new keynesian perspective. *Journal of Economic Literature*, 37(4):1661–1707, 1999.

Marco Di Maggio, Amir Kermani, and Christopher J. Palmer. How quantitative easing works: Evidence on the refinancing channel. *The Review of Economic Studies*, 87(3): 1498–1528, 2020.

Christopher J. Erceg, Dale W. Henderson, and Andrew T. Levin. Optimal monetary policy with staggered wage and price contracts. *Journal of Monetary Economics*, 46(2):281–313, 2000.

European Insurance and Occupational Pensions Authority. Artificial intelligence governance principles: Towards ethical and trustworthy artificial intelligence in the european insurance sector. Technical report, EIOPA, 2021.

Antonio Falato, Dalida Kadyrzhanova, Jae Sim, and Roberto Steri. Rising intangible capital, shrinking debt capacity, and the us corporate savings glut. *The Journal of Finance*, 77(5):2799–2852, 2022.

Financial Stability Board. Artificial intelligence and machine learning in financial services. Technical report, FSB, 2017.

Financial Stability Board. The financial stability implications of artificial intelligence. Technical report, Financial Stability Board, 2024.

Quirin Fleckenstein, Manasa Gopal, Germán Gutiérrez, and Sebastian Hillenbrand. Nonbank lending and credit cyclicality. *The Review of Financial Studies*, page hhaf024, 2025.

Anne Fournier, Ralf Meisenzahl, and Andy Polacek. Privately placed debt on life insurers' balance sheets: Part 1—a primer. Chicago Fed Letter 493, Federal Reserve Bank of Chicago, May 2024.

Fulvia Fringuellotti and João A. C. Santos. Insurance companies and the growth of corporate loans' securitization. *FRB of New York Staff Report*, 975, 2025.

Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. Predictably unequal? the effects of machine learning on credit markets. *Journal of Finance*, 77(1):5–47, 2022.

Luca Gagliardone, Mark Gertler, Simone Lenzu, and Joris Tielens. Micro and macro cost-price dynamics in normal times and during inflation surges. Technical report, National Bureau of Economic Research, 2025a.

Luca Gagliardone, Mark Gertler, Simone Lenzu, and Joris Tielens. Anatomy of the phillips curve: micro evidence and macro implications. *American Economic Review*, 115(11): 3941–3974, 2025b.

Jordi Galí. *Monetary policy, inflation, and the business cycle: an introduction to the new Keynesian framework and its applications*. Princeton University Press, 2015.

Gary Gensler. Isaac newton to ai. Technical report, U.S. Securities and Exchange Commission, 2023. Speech at the National Press Club.

Mark Gertler and Peter Karadi. A model of unconventional monetary policy. *Journal of Monetary Economics*, 58(1):17–34, 2011.

Mark Gertler and John Leahy. A phillips curve with an ss foundation. *Journal of Political Economy*, 116(3):533–572, 2008.

Mark Gertler and Antonella Trigari. Unemployment fluctuations with staggered nash wage bargaining. *Journal of Political Economy*, 117(1):38–86, 2009.

Gita Gopinath. Statement by imf first deputy managing director gita gopinath at the conclusion of the third meeting of g20 finance ministers and central bank governors. Technical report, International Monetary Fund, July 2025.

Kathryn Holston, Thomas Laubach, and John C. Williams. Measuring the natural rate of interest: International trends and determinants. *Journal of International Economics*, 108:S39–S75, 2017.

Chang-Tai Hsieh and Peter J Klenow. Misallocation and manufacturing tfp in china and india. *The Quarterly journal of economics*, 124(4):1403–1448, 2009.

International Monetary Fund. Global financial stability report, october 2024: Chapter 3 – advances in artificial intelligence and financial stability. Technical report, International Monetary Fund, 2024.

Victoria Ivashina. Private credit: What do we know? Technical report, Harvard Business School and NBER, October 2025. Working paper.

Young Soo Jang, Dasol Kim, and Amir Sufi. The lending technology of direct lenders in private credit. Technical report, Working Paper, November 2025.

Sophia Kazinnik and Erik Brynjolfsson. Ai and the fed. Technical report, National Bureau of Economic Research, 2025.

Miles S Kimball. The quantitative analytics of the basic neomonetarist model. *Journal of Money, Credit, and Banking*, 27(4):1241–1277, 1995.

Andrei Kirilenko, Albert Kyle, Mehrdad Samadi, and Tugkan Tuzun. The flash crash: High-frequency trading in an electronic market. *Journal of Finance*, 72(3):967–998, 2017.

Thomas Laubach and John C. Williams. Measuring the natural rate of interest. *Review of Economics and Statistics*, 85(4):1063–1070, 2003.

Simone Lenzu, Filippo Mezzanotti, and Joris Tielens. A early assessment of the adoption, use, and impact of genai on firms' production, productivity, and profitability. 2026.

Christopher McMahon, Donald McGillivray, Ajit Desai, Francisco Rivadeneyra, Jean-Paul Lam, Thomas Lo, Danica Marsden, and Vladimir Skavysh. Improving the efficiency of payments systems using quantum computing. *Management Science*, 70(10):7325–7341, 2024.

Emi Nakamura and Jón Steinsson. Monetary non-neutrality in a multisector menu cost model. *The Quarterly Journal of Economics*, 125(3):961–1013, 2010.

OECD/BCG/INSEAD. The adoption of artificial intelligence in firms: New evidence for policymaking, 2025.

The Economist. What if the $3trn ai investment boom goes wrong? *The Economist*, September 2025.

Olivier Wang and Iván Werning. Dynamic oligopoly and price stickiness. *American Economic Review*, 112(8):2815–49, 2022.

Michael Woodford. *Interest and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, 2011.

A Model appendix

A.1 Setup

The economy is populated by heterogeneous producers (or firms), denoted by f, each operating in an industry $i \in \mathcal{I} = [0, 1]$. We denote by $\mathcal{F}_i$ the set of firms competing in industry i. Each firm is measure zero relative to the economy as a whole but may be large relative to its industry. Hence, it takes the aggregate expenditure as given but internalizes the effect of its pricing decisions on the consumption and price index of its industry. Time is discrete.

Let P_{ft} denote the price charged by each firm for a unit of its output, P_{it} the industry price index, φ_{ft} a firm-specific relative demand shifter, and Y_{it} the real industry output. For any industry i, we consider an arbitrary, invertible demand system that generates a residual demand function of the following form:

$$\mathcal{D}_{ft} := d(P_{ft}, P_{it}, \varphi_{ft})Y_{it} \qquad \forall f \in \mathcal{F}_i. \tag{A.1}$$

Below we follow closely the steps in Gagliardone et al. (2025b) to characterize the firms' pricing problem and derive the cost-based NKPC. We extend the framework by allowing for a richer characterization of real marginal cost which accounts for cyclical variation in factor market frictions and technology.

A.2 The firm pricing problem

Firms adjust their prices during each period in order to maximize expected profits facing nominal rigidities as in Calvo (1983). Each period they face a probability $(1 - \theta) \in [0, 1]$ of being able to change their price, independent across time and across firms. Thus, the price P_{ft} paid by consumers to buy goods produced by firm f is either the optimal reset price if the firm is able to adjust, denoted by P^o_{ft}, or the price it charged in the previous period, P_{ft-1}.

When choosing P^o_{ft}, firms consider both their own costs, the pricing choices made by competitors, as well as the impact of their own price adjustments on their residual demand and on the industry-wide price index. Let $\Lambda_{t,\tau}$ denote the stochastic discount factor between time t and $t + \tau$, $TC_{ft} := TC(\mathcal{D}_{ft})$ the real total costs, and MC^n_{ft} the nominal marginal cost of firm f (which we characterize below). Then the optimal reset price P^o_{ft} solves the following profit maximization problem:

$$\max_{P^o_{ft},\{Y_{ft+\tau}\}_{\tau\geq 0}} \mathbb{E}_t \left\{ \sum_{\tau=0}^{\infty} \theta^\tau \left[\Lambda_{t,\tau} \left(\frac{P^o_{ft}}{P_{t+\tau}} \mathcal{D}_{ft+\tau} - TC(\ \mathcal{D}_{ft+\tau}) \right) \right] \right\},$$

subject to the sequence of expected demand functions $\{\mathcal{D}_{ft+\tau}\}_{\tau\geq 0}$ in Equation (A.1). Nominal rigidities generate forward-looking pricing behavior, as firms take into account that it might not be possible to adjust prices every period. As a result, the optimal reset price is a weighted average of current and expected future nominal marginal costs and markups. Denoting by μ_{ft} the desired log markup, the FOC of the problem is:

$$\mathbb{E}_t \left\{ \sum_{\tau=0}^{\infty} \theta^\tau \Lambda_{t,\tau} \mathcal{D}_{ft+\tau} \left[\frac{P^o_{ft}}{P_{t+\tau}} - (1+\mu_{ft+\tau}) \frac{MC^n_{ft+\tau}}{P_{t+\tau}} \right] \right\} = 0. \tag{A.2}$$

Thus, the optimal reset price depends on the expected path of marginal cost and desired markups over the period the firm expects its price to be fixed, where θ^τ is the probability the firm expects its price to be fixed τ periods from now.

We log-linearize the FOC in Equation (A.2) around the symmetric steady state with zero inflation.[19] Denoting the variables in logs with lower-case letters, we obtain that the reset price satisfies:

$$p^o_{ft} = (1-\beta\theta)\mathbb{E}_t \left\{ \sum_{\tau=0}^{\infty} (\beta\theta)^\tau \left(\mu_{ft+\tau} + mc^n_{ft+\tau} \right) \right\}. \tag{A.3}$$

The log-linearized desired markup (in deviation from steady state markup μ_f) is a function that depends inversely on the log-difference between the firms' own reset price and its competitors' prices (p^{-f}_{it}):

$$\mu_{ft} - \mu_f = -\Gamma \left(p^o_{ft} - p^{-f}_{it} \right) + u^\mu_{ft}, \tag{A.4}$$

where $\Gamma > 0$ denotes the markup elasticity with respect to prices and u^μ_{ft} is a firm-specific demand shock to the desired markup that depends on the demand shifter φ_{ft}. Gagliardone et al. (2025b) show that, under weak assumptions, the expression in Equation (A.4) holds for standard frameworks with imperfectly competitive firms, including monopolistic competition with variable elasticity of demand (Kimball 1995), static oligopoly (Atkeson and Burstein 2008) and dynamic oligopoly (Wang and Werning 2022). These frameworks share the property that, in equilibrium, a firm's elasticity of demand declines as its market

[19] The choice of the zero-inflation steady state permits simpler notation; but is largely immaterial for our purposes.

share increases. Thus, the presence of strategic complementarities in price setting implies that a relative price increase lowers a firm's desired markup, dampening the response of prices to marginal cost.

Substituting the expression for $\mu_{ft+\tau}$ in the log-linearized first-order condition, we obtain the following forward-looking pricing equation:

$$p^o_{ft} = (1-\beta\theta)\Theta\mathbb{E}_t\left\{\sum_{\tau=0}^{\infty}(\beta\theta)^\tau\left((1-\Omega)(mc^n_{ft+\tau}+\mu_f)+\Omega p^{-f}_{it+\tau}\right)\right\}+u_{ft}, \tag{A.5}$$

where u_{ft} captures residual variation in the markup that depends on the aggregation of firms' demand shifters and the changes in the slope of competitors' reaction function. For the purposes of this paper, we ignore this term and set it to zero.

The parameter $\Omega := \frac{\Gamma}{1+\Gamma}$ captures the strength of strategic complementarities and impacts the firm's pricing policy by muting the price response to changes in marginal costs. If the demand elasticity is constant, as in the textbook New Keynesian model with monopolistically competitive firms, the desired markup is a constant. In this case, $\Omega = 0$ and the optimal pricing equation simplifies to the familiar formulation where the reset price exclusively depends on the current and future stream of marginal costs. Competitors' prices are then irrelevant.

The parameter $\Theta \leq 1$ captures macroeconomic complementarities due to aggregate returns to scale in production. For example, under CES demand with elasticity of substitution γ, we have that $\Theta := \frac{1}{1+\gamma(1-\alpha)(1-\Omega)}$. A higher elasticity of substitution increases competitive pressure and magnifies the aggregate response of costs. Similarly, a higher elasticity of cost to output—reflecting stronger decreasing returns or tighter capacity constraints—amplifies aggregate marginal-cost pressures as output expands, strengthening macroeconomic complementarities and lowering Θ. Strategic complementarities interact with these forces in a subtle way. A higher degree of complementarities in price setting weakens macroeconomic complementarities by dampening the amplification of marginal costs through output.

A.3 Aggregation and the cost-based New Keynesian Phillips curve

To obtain closed form expressions, suppose there are $N < \infty$ in each industry i competing a la Bertrand, and order firms in each industry from 1 to N.[20] The aggregate price index (in log-linear terms) is:

$$p_t = \int_{i\in I}\left(\frac{1}{N}\sum_{f=1}^{N} p_{fit}\right)di,$$

(In the paper, we dropped the industry subscript for ease of notation.) Denote by $B^{\star}_{ft}$ for $f \in \{1,\ldots,N\}$ the set of industries in which the f-th firm can adjust. The price index can then be rewritten as:

$$p_t = \frac{1}{N}\sum_{f=1}^{N}\left(\int_{i\in I/A^{\star}_{ft}} p_{fit-1}di + \int_{i\in A^{\star}_{ft}} p^{o}_{fit}di\right),$$

where we are using the fact that firms that cannot adjust set their price to their $t-1$ level, whereas firms that can adjust set it to the optimal reset price.

Since $B^{\star}_{ft}$ has measure $1-\theta$, and the identity of firms that adjust is an i.i.d. draw from the total population of firms, using the law of large numbers for each $f = \{1,\ldots,N\}$ across industries we have that:[21]

$$\frac{1}{N}\sum_{f=1}^{N}\int_{i\in I/B^{\star}_{ft}} p_{fit-1}di = \theta\int_{i\in I}\left(\frac{1}{N}\sum_{f=1}^{N}p_{fit-1}\right)di = \theta p_{t-1}$$

and

$$\frac{1}{N}\sum_{f=1}^{N}\int_{i\in B^{\star}_{ft}} p^{o}_{fit}di = (1-\theta)\int_{i\in I}\left(\frac{1}{N}\sum_{f=1}^{N}p^{o}_{fit}\right)di.$$

Defining the average reset price in the economy:

$$p^{o}_t := \int_{i\in I}\left(\frac{1}{N}\sum_{f=1}^{N}p^{o}_{fit}\right)di,$$

we obtain an equation characterizing the log-linear aggregate price index:

$$p_t = (1-\theta)p^{o}_t + \theta p_{t-1}, \tag{A.6}$$

[20] Letting $N \to \infty$, all results hold under Kimball preferences. Note also that the same argument goes through with minor modifications, but heavier notation, for $N_i \neq N$ for a non-zero measure of industries.

[21] The i.i.d. assumption implies that: $\int_{i\in B\subseteq[0,1]} p_{fit}di = Pr(B)\int_{i\in I} p_{fit}di$. Notice also that $\int_{i\in[0,1]}\left(\frac{1}{N}\sum_{f=1}^{N}p^{-f}_{it}\right)di = \int_{i\in[0,1]}\left(\frac{1}{N}\sum_{f=1}^{N}\left[\frac{N}{N-1}p_{it} - \frac{1}{N-1}p_{fit}\right]\right)di = p_t$.

with p_t and p_t^o denoting the aggregate price indices implied by the demand system. Next, we replace the aggregate reset price, p_t^o, with an expression that depends on aggregate marginal costs and prices.

Let $mc_t = mc_t^n - p_t$ denote aggregate real marginal cost (characterized below) and define aggregate inflation as $\pi_t = p_t - p_{t-1}$. Following the steps in Gagliardone et al. (2025b), we average across firms and industries to obtain an expression for the aggregate reset price:

$$p_t^o = (1-\beta\theta)\left((1-\Omega)\Theta\widehat{mc}_t + p_t\right) + \beta\theta\mathbb{E}_t p_{t+1}^o$$

Subtracting p_t from both sides and using the log-linearized price index:

$$p_t^o - p_t = (1-\beta\theta)(1-\Omega)\Theta\widehat{mc}_t + \beta\theta(\mathbb{E}_t p_{t+1}^o - p_t)$$

Rearranging and combining the equation for the log-linear price aggregate price index in A.6 with the equation above gives the primitive formulation of the NKPC curve:

$$\pi_t = \lambda\, \widehat{mc}_t + \beta\, \mathbb{E}_t\{\pi_{t+1}\}, \tag{A.7}$$

with the slope given by:

$$\lambda := \frac{(1-\theta)(1-\beta\theta)}{\theta}(1-\Omega)\Theta. \tag{A.8}$$

A.4 Derivation of the real marginal cost gap $\widehat{mc}_t$

To derive the aggregate real marginal cost gap we start from the derivation of real marginal cost at the firm-level, $mc_{fi,t}$, and aggregate by averaging across firms and industries. In doing so, we omit the firm and industry subscript (f, i) for ease of notation.

Each firm chooses a bundle of variable inputs $X_{j,t}$, $j = 1, \ldots, J$ (e.g., different types of labor, intermediate inputs, or capital services), with nominal user costs collected in the vector $W_t^n = (W_{1,t}^n, \ldots, W_{J,t}^n)$. Output is produced according to the technology

$$Y_t = A_t\, H(X_t)^{\alpha},$$

where A_t is Hicks-neutral productivity, $H(\cdot)$ is an aggregate-input index, and $\alpha > 0$ governs returns to scale in the mapping from aggregate inputs to output.

Throughout, a superscript $\star$ denotes the flexible-price benchmark allocation, holding fixed the real frictions that characterize the natural allocation. Hence, real wedges need not vanish in the flexible-price equilibrium. Lowercase letters denote natural logarithms, and for any variable z_t we define $\widehat{z}_t \equiv z_t - z_t^{\star}$.

Cost function and nominal unit cost. Define the nominal cost function

$$C(Y_t, A_t, W_t^n, \Xi_t) \;\equiv\; \min_{X_t} \sum_{j=1}^{J} W_{j,t}^n X_{j,t} \quad \text{s.t.} \quad Y_t \le A_t \big(H(X_t)\big)^{\alpha}.$$

The term Ξ_t is a reduced-form placeholder for real input-market frictions (e.g., bargaining wedges, markups embedded in intermediate-input prices, or financing premia embedded in user costs). These frictions are not assumed away under flexible prices, so $\Xi_t^{\star}$ is generally nonzero.

Assume that the aggregate-input index $H(\cdot)$ is homogeneous of degree one. Then producing Y_t units of output requires an aggregate input level satisfying $H(X_t) \;\ge\; (Y_t/A_t)^{1/\alpha}$. By homogeneity, the cost-minimization problem admits a two-stage representation:

$$C(Y_t, A_t, W_t^n, \Xi_t) = \left(\frac{Y_t}{A_t}\right)^{1/\alpha} \cdot \Gamma(W_t^n, \Xi_t),$$

where $\Gamma(W_t^n, \Xi_t)$ is the minimum nominal cost of producing one unit of the aggregate input H.

Nominal marginal cost. Nominal marginal cost is defined as:

$$MC_t^n \;\equiv\; \frac{\partial C(Y_t, A_t, W_t^n, \Xi_t)}{\partial Y_t}.$$

Differentiating with respect to Y_t yields:

$$MC_t^n = \Gamma(W_t^n, \Xi_t) \cdot \frac{\partial}{\partial Y_t}\left(\frac{Y_t}{A_t}\right)^{1/\alpha} = \Gamma(W_t^n, \Xi_t) \cdot \frac{1}{\alpha}\left(\frac{Y_t}{A_t}\right)^{1/\alpha} \frac{1}{Y_t}.$$

Equivalently,

$$MC_t^n = \frac{1}{\alpha}\, \Gamma(W_t^n, \Xi_t)\, Y_t^{\frac{1}{\alpha}-1}\, A_t^{-\frac{1}{\alpha}}.$$

It is convenient to define $\chi \equiv (1/\alpha) - 1$, so that $\chi > 0$ corresponds to decreasing returns to scale ($\alpha < 1$), while $\chi < 0$ corresponds to increasing returns to scale ($\alpha > 1$). We then have that:

$$MC_t^n = \frac{1}{\alpha}\, \Gamma(W_t^n, \Xi_t)\, Y_t^{\chi}\, A_t^{-(1+\chi)}.$$

Taking logs, we obtain

$$mc_t^n \;\equiv\; \log MC_t^n = \log \Gamma(W_t^n, \Xi_t) + \chi\, y_t - (1+\chi)\, a_t^{tfp} - \log \alpha,$$

where $a_t^{tfp} \equiv \log A_t$ denotes log technical efficiency.

Real user-cost index and wedges. We decompose the real unit cost implied by the cost function into a component reflecting equilibrium factor prices and wedges:

$$\log\Gamma(W_t^n, \Xi_t) - p_t = \left(\log\Gamma(W_t^n, 0) - p_t\right) + \tau_t = (w_t^n - p_t) + \tau_t,$$

where $w_t^n \equiv \log\Gamma(W_t^n, 0)$ denotes the (optimized) nominal unit cost implied by factor prices in the absence of the additional real frictions; τ_t captures distortions summarized by Ξ_t, that shift effective user costs independently of equilibrium factor prices. Such wedges may reflect, among others, bargaining frictions in labor markets, financing premia embedded in user costs, markups in intermediate-input prices, or other input-market distortions. Importantly, τ_t represents an aggregate (or average) factor-market wedge in the spirit of Chari et al. (2007). These aggregate wedges are conceptually distinct from misallocation wedges arising from cross-sectional heterogeneity in the frictions faced by individual firms when accessing factor markets (Hsieh and Klenow 2009; Baqaee et al. 2024). We capture such misallocation effects instead through the productivity disturbance $\tilde{a}_t$, defined below.

Up to first-order, we can express the user-cost index w_t^n and the factor market's wedge τ_t as a function of of individual factors' prices and wedges. To see this, recall that the nominal cost function admits the representation $C(Y_t, A_t, W_t^n, \Xi_t) = (Y_t/A_t)^{1/\alpha}\,\Gamma(W_t^n, \Xi_t)$. Because the scale term $(Y_t/A_t)^{1/\alpha}$ does not depend on individual input prices, Shephard's lemma applies directly to $\Gamma(W_t^n, \Xi_t)$, the minimum nominal cost of producing one unit of the aggregate input $H(\cdot)$. Let $c_{j,t}$ denote the cost share of input j:

$$c_{j,t} \;\equiv\; \frac{W_{j,t}^n X_{j,t}}{C(Y_t, A_t, W_t^n, \Xi_t)} = \frac{\partial \log\Gamma(W_t^n, \Xi_t)}{\partial \log W_{j,t}^n}.$$

Evaluating this expression at the flexible-price benchmark yields the reference cost shares $c_j^\star$, which are time-invariant to a first-order approximation because they are computed at the natural allocation. Taking a first-order Taylor expansion of $\log\Gamma(W_t^n, \Xi_t)$ around the flexible-price allocation $(W_t^{n\star}, \Xi_t^\star)$ and subtracting $(p_t - p_t^\star)$, we obtain the first-order approximation:

$$\widehat{w}_t \;\equiv\; w_t - w_t^\star \;\approx\; \sum_{j=1}^{J} c_j^\star\,\widehat{w}_{j,t}, \qquad \widehat{\tau}_t \;\equiv\; \tau_t - \tau_t^\star \;\approx\; \sum_{j=1}^{J} c_j^\star\,\widehat{\tau}_{j,t},$$

This decomposition clarifies the distinction between real equilibrium factor-price pressures, summarized by $\widehat{w}_t$, and cyclical distortions in input markets that shift marginal costs independently of factor prices, summarized by $\widehat{\tau}_t$, both of which enter the marginal-cost gap that drives inflation dynamics.

Real marginal cost. Define real marginal cost as nominal marginal cost divided by the aggregate price level. In logs, $mc_t \equiv mc_t^n - p_t$. Substituting the expressions above yields:

$$mc_t = w_t + \tau_t + \chi\, y_t - a_t - \log \alpha,$$

where we define the *real unit cost index* as $w_t \equiv w_t^n - p_t$ and *effective productivity*—the component of efficiency relevant for marginal cost—as:

$$a_t \equiv (1 + \chi)\, a_t^{tfp} + \tilde{a}_t.$$

The term $\tilde{a}_t$ captures cyclical distortions in technical efficiency arising from price dispersion, wage dispersion, misallocation of inputs (Hsieh and Klenow 2009; Baqaee et al. 2024), or congestion in factor markets. Under flexible prices, real marginal cost is $mc_t^\star = w_t^\star + \tau_t^\star + \chi\, y_t^\star - a_t^\star - \log \alpha$, where by construction, the productivity distortions vanish in the flexible-price allocation, so that $\tilde{a}_t^\star = 0$.

Define the *cost pressure index* $q_t \equiv \chi\, y_t + w_t + \tau_t$. Subtracting term by term the flexible price benchmark from the time corresponding t , we obtain decomposition of the primitive NKPC in terms of the real marginal cost gap $\widehat{mc}_t$ used in Equation (1) the paper:

$$\begin{aligned} \widehat{mc}_t &\equiv mc_t - mc_t^\star \\ &= \widehat{q}_t - \widehat{a}_t \\ &= \chi\, \widehat{y}_t + \widehat{w}_t + \widehat{\tau}_t - \widehat{a}_t \end{aligned} \tag{A.9}$$

This expression shows that deviations of real marginal cost from its flexible-price benchmark reflect four distinct forces: scale effects associated with deviations of output from potential ($\chi\, \widehat{y}_t$), cyclical movements in real input prices ($\widehat{w}_t$), cyclical input-market wedges ($\widehat{\tau}_t$), and deviations in effective productivity relative to the flexible-price allocation ($\widehat{a}_t$). The latter captures efficiency losses due to misallocation and dispersion induced by nominal rigidities.

Aggregation. Aggregating across firms and industries, we obtain the aggregate (aka, average) real marginal cost gap that enters the primitive NKPC as a forcing variable.

$$\widehat{mc}_t \equiv \int_i \sum_{f=1}^{N} \frac{1}{N} \sum_{f=1}^{N} mc_{fit} \, di - \int_i \sum_{f=1}^{N} \frac{1}{N} \sum_{f=1}^{N} mc^{\star}_{fit} \, di$$

Recall that we assumed that each firm faces the same input prices and that the wedges τ_t are aggregate wedges, which implies that t $w_{fit} = w_t$ and $\tau_{fit} = \tau_t$. Thus the construction of aggregate real marginal cost requires averaging only across firm's output and and realized productivity (y_{fit}, a_{fit}).