Artificial Intelligence, Competition, and Welfare

Susan Athey and Fiona Scott Morton[1]

19 September 2025

Abstract

We propose a policy-relevant research agenda examining how market power in upstream artificial intelligence (AI) affects downstream prices, industry structure, factor returns, and welfare—especially whether labor-displacing AI leaves workers worse off. In our open-economy general equilibrium model, AI is a priced, imported input. Distributional effects depend on how sectoral skill intensity responds to AI prices. Non-monotonicity can cause "double harm" for displaced workers, who may face lower real wages both when AI is cheap and again if prices rise due to market power. Our main model features two non-traded sectors and firms making discrete adoption decisions about technology. Adoption reduces unit costs, displaces some types of workers, and depresses wages for those workers via diminishing returns elsewhere, while leaking AI fees abroad. Strategic AI pricing reduces welfare by raising downstream marginal costs (via usage fees) and limiting entry and variety (via access fees). We derive an adoption frontier linking feasible usage fees to displaced workers' outside options, showing that a monopolist typically prices on this boundary; capping one fee shifts rents to the other. Regulating both fees, alongside policies that absorb displaced labor, can raise national welfare.

[1] We are grateful to our discussant, Catherine Tucker, as well as conference participants and particularly Daron Acemoglu, Joshua Gans, Sam Kortum, Stephen Redding, Pascuale Restrepo, Joseph Stiglitz, Dan Treffler, Ellie Tyger, for insightful comments. Nick Jacobson provided careful research assistance. All errors are our own.

Introduction

Economists have observed that artificial intelligence (AI) may have a variety of downstream benefits for the economy: it may improve productivity, enable innovation and increase entry of new firms. Alongside potential benefits, AI poses economic risks. Chief among these are worker displacement and industry restructuring, with consequences for income distribution both within and across countries.

In this paper, we focus on what we view as a critical factor affecting the magnitudes and balance of benefits and risks: market power in AI. We propose a research agenda that studies the impact of such market power on downstream industries, including impacts on downstream industry structure, profits, innovation, and consumer prices. We further highlight indirect effects on wages, inequality, productivity in other sectors, and welfare. When the profits obtained by the AI provider are not recirculated in the economy, for example, if the country imports AI products, welfare in a country may fall in aggregate as a result of the introduction of AI, and specific groups may be harmed by falling wages without corresponding increases in the variety of goods or decreases in prices.

Competition authorities worldwide have recognized that AI may create new bottlenecks, leading to higher prices, reduced innovation, and lower benefits for consumers. The UK's Competition and Markets Authority, the US FTC, and other agencies have issued reports and are monitoring commercial relationships within the AI stack.[2] Scrutiny may be warranted because different layers of the AI stack—from chips to foundation models to applications and distribution channels—exhibit concentration. Examples include NVIDIA's dominance in advanced chips, concentration among foundation model providers, and the dependence of downstream applications (e.g. search engines, language tutors, writing assistants, customer service platforms) on a few foundation models. Access to data that is crucial for AI performance may be gated by incumbents in both the consumer and enterprise software markets. Distribution channels, such as mobile operating systems, productivity software, and online platforms, may become bottlenecks given that many are controlled by incumbents with substantial market power. When these channels are controlled by firms that also supply AI, or by firms that have control over critical proprietary data, entrants face foreclosure risk, and even efficient AI applications may struggle to gain distribution, data and scale.

In this paper, we focus on the business-to-business use case, where AI is an input to large parts of the economy. We assume AI is imported, which may be accurate for countries that do not participate in the AI value chain, and it may be a useful approximation for scenarios where domestic firm profits are not broadly recirculated in the domestic economy. We argue that the extent to which AI providers sell at competitive prices has critical implications for the impact of AI on income distribution and welfare. We consider

---

[2] see, e.g., Competition and Markets Authority (2023; Portuguese Competition Authority 2023).

extensions and variants of standard general equilibrium and trade models. Unlike much of the literature that focuses primarily on distribution, we consider assumptions that are rich enough such that productivity improvements from AI do not necessarily result in improvements in economy-wide welfare, let alone welfare for specific groups.

Our starting point in this paper is to observe that the arrival of a new technology can be modeled similarly to the availability of a new input factor available through trade. Thus, we analyze the impact of AI using models of a small economy, considering cases of open and closed economies. However, we include a competition element in the model by treating AI as an intermediate input X supplied by a foreign monopolist or supplied competitively. The price of AI is therefore not an exogenous world price, but a strategic variable chosen by the upstream supplier, where the supplier may set the price to maximize revenue, but may also be regulated. The goal is to understand how the introduction of AI and subsequent changes in its price redistribute income across factors, shift output, and affect aggregate welfare. We are particularly interested in the real wages of workers, and whether high prices for AI lead to reductions in the real wage despite efficiency benefits in production. In extensions, we highlight the particularly challenging but realistic scenario where AI providers can use nonlinear pricing schemes, influencing entry, innovation and industry structure in downstream industries.

In our models the supplier of AI may be able to extract the surplus created by AI technology, while, at the same time, the exercise of market power inflicts additional loss on the workers displaced by the AI. This contrasts with a standard model where, if a technology is introduced that disproportionately displaces a group of workers, raising prices for the input helps that group by diminishing the substitution.

We also highlight the key question of whether other sectors, either nontraded sectors or export sectors, can productively absorb displaced workers, an assumption that may be questionable in the face of a major technological shift. This focus underscores the critical role played by government policies that change how productive workers are in sectors *less* affected by AI. For example, the government may increase spending on services like healthcare and education and then help displaced workers transition into those industries, perhaps using AI to be more productive in their new roles.

The popular interest in AI has stimulated a macroeconomic literature modeling possible faster growth from AI. Aghion et al. (2019) incorporate AI into a growth model to explore implications for the labor–capital ratio and the share of expenditure on automated tasks. Outcomes depend on parameters such as substitution elasticities between automated and non-automated goods, the speed of cost reductions for automated goods, and the proportion of tasks that can be automated. Nordhaus (2021) analyzes the channels through which AI can increase economic growth. He presents a model in which AI induces capital deepening, which in turn accelerates growth. Nordhaus emphasizes that returns accrue primarily to capital: *"Capital eventually gets virtually all the cake, but the crumbs left for labor—which are really small pieces of the increasingly huge mountains of cake—are still growing at a phenomenal rate"* (p. 14). In our setting, the question is whether labor actually does benefit from the crumbs, if the things capital produces remain expensive to them.

Studies of task–based and automation frameworks (for an early example, see (Autor et al. 2003)) study how technology displaces or complements labor when technology costs and output prices are taken as parametric. Growth papers model AI as a driver of factor usage and total factor productivity (e.g., (Aghion et al. 2019; Nordhaus 2021)); open-economy work on directed technical change treats final goods as traded and incidence of impacts as disciplined by terms-of-trade movements (e.g., (Korinek and Stiglitz 2018, 2021)). We incorporate these forces in our model.

(Korinek and Stiglitz 2018, 2021) propose a series of models that focus on distributional consequences and policy responses from exogenous technological shocks or changes in the returns to the resources of small countries. A complementary line of work, going back to Stiglitz (1976) and further developed in Delli Gatti et al. (2012; Gatti et al. 2012), studies dual-economy settings in which a constant-returns agricultural sector coexists with an urban sector featuring wage rigidity (e.g., due to efficiency-wage considerations). In their baseline environment, an agricultural productivity improvement is *unambiguously* welfare-reducing: higher rural productivity contracts urban employment when nominal wages cannot adjust, and with flexible wages an induced wage decline can further depress demand and raise unemployment. Our models connect to these results but identify distinct channels that do not rely on nominal rigidities.

Our framework also connects to two classic ideas in economics—Baumol's "cost disease" and the "Dutch disease"—but also shows why the forces they highlight may not apply or may be more nuanced in the case of AI. More precisely, classic Dutch disease logic (Corden–Neary, 1982) raises nontraded prices when resources are pulled into a booming sector, and Baumol–Bowen (1966) shows that uneven productivity growth can raise relative prices in stagnant sectors because wages equalize economy-wide. With AI, we argue that the analogy is incomplete. Adoption of AI often *releases* labor rather than absorbing it, so there is little resource pull into the "booming" activity; displaced workers crowd into nontraded services, and whether their prices rise depends on wages and sectoral productivities (identified by Baumol) rather than on reallocation alone. We formalize these benchmarks, and build up to our main model that adds entry/variety and analyzes the welfare impact of market power and nonlinear pricing of AI.

A related perspective comes from growth accounting. Domar (1961) and Hulten (1978) showed that the aggregate effect of a sector's productivity change depends on its "Domar weight"—the ratio of its gross output to GDP. Because an upstream general purpose technology supplies inputs to many final sectors, its Domar weight is disproportionately large and shocks there propagate strongly through the economy. In our setting, AI plays exactly this role: changes in its usage fee resemble productivity shocks in a sector with very high Domar weight. As Baqaee and Farhi (2019, 2020) emphasize, however, the Hulten–Domar formula captures only first-order effects. General equilibrium forces can overturn the first-order gain. In our framework, three forces are central: (i) **labor**

displacement, which lowers wages in the sectors that absorb redundant workers; (ii) **nontraded scarcity**, which raises the cost of living if national income rises; and (iii) **market power**, which allows a monopolist AI supplier to charge high usage and access fees, leak rents abroad, and blunt pass-through of productivity gains. Our contribution is to highlight the third channel. Market power shapes outcomes: if rents are retained domestically and income rises, nontraded prices climb and the cost of living worsens; if rents leak abroad and income stagnates, nontraded prices may not increase, but displaced workers are still worse off because real wages fall and national income is lower. Taken together, these forces imply that even if AI appears beneficial in a Hulten–Domar sense, cheaper AI can still reduce real wages and national welfare once displacement, nontraded scarcity, and monopoly rents are accounted for.

We propose a general equilibrium model where AI is a *priced, imported factor* whose upstream market power is central. This allows us to separate (i) pure technological efficiency from (ii) the impact of the *price path of AI services* and the corresponding changes in factor markets and input markets, and (iii) to ask how welfare depends jointly on whether AI prices fall with productivity and on whether displaced workers are sufficiently productive in sectors less affected by AI. In our model, the firms adopting AI do not internalize the negative externality on worker wages, and they do not internalize that AI payments leave the economy rather than flowing back into country income. We show that conclusions about wages and aggregate welfare turn on a small set of modeling choices that are especially salient for AI: whether goods are traded versus nontraded (and in turn, whether there is diminishing marginal utility in the output markets); whether automation reduces demand for some or all groups of labor to (near) zero so that output expansions in newly productive industries do not lead to increases in labor demand; and whether the upstream supplier can use nonlinear pricing.

Throughout, the *price of AI*, denoted p_X, is a strategic object, one which reallocates income between factors and shifts national income. We consider both continuous and discrete production functions, where a large fall in the price of an AI input may fundamentally change which factor of production (e.g. skilled or unskilled labor) is dominant, and where labor demand may be insensitive to input prices for technologies that fully automate. We also pay special attention to the choices firms make about adopting AI in the face of a strategic monopolist setting prices for the AI input to extract surplus.

Benchmarks: Incorporating AI in Standard Models.

We begin by highlighting forces that arise in the standard two–good Heckscher–Ohlin model with AI as an imported factor; we further show how, with larger technical shifts that may occur with AI, income inequality can experience what we call “double-harm.” The details of the analysis are developed in the Online Appendix. In the first baseline, sector A produces using skilled and unskilled labor (L_S, L_U) alone; sector B combines (L_S, L_U) with AI, purchased at price p_X from the foreign supplier. We write the quantity of labor type $i \in$

$\{S, U\}$ in sector $j \in \{A, B\}$ as i_j. Both goods A and B are traded at exogenous world prices $(\bar{p}_A, \bar{p}_B)$.

When $p_X = \infty$, AI is unavailable. When p_X is finite, sector B's costs c_B depend on (w_S, w_U, p_X), where (w_S, w_U) are the equilibrium wages of skilled and unskilled labor. Throughout, we measure the AI input price p_X in units of the price of good A.

Zero–profit conditions determine (w_S, w_U) as a function of p_X. For each sector $j \in \{A, B\}$, let $\theta_i^j = w_i a_i^j / c_j$ denote the cost share of factor $i \in \{S, U\}$, and let $R_j = \theta_S^j / \theta_U^j$ denote the sectoral intensity ratio. B is skill–intensive relative to A if $R_B > R_A$, and unskilled–intensive if $R_B < R_A$.

For expositional simplicity, we assume that for sufficiently high p_X, sector B's skill intensity $R_B(p_X) := \theta_S^B / \theta_U^B$ is *decreasing* in p_X (equivalently, *increasing* as AI gets cheaper).

General-equilibrium mapping and reallocation.

Totally differentiating zero-profit in A and B at fixed $(\bar{p}_A, \bar{p}_B)$ yields

$$\theta_S^A \, dw_S + \theta_U^A \, dw_U = 0, \qquad \theta_S^B \, dw_S + \theta_U^B \, dw_U + \theta_X^B \, dp_X = 0.$$

Eliminating dw_U,

$$\begin{aligned} &(\theta_S^B - \theta_U^B \, \theta_S^A / \theta_U^A) \, dw_S = -\, \theta_X^B \, dp_X, \\ &\Rightarrow \quad \mathrm{sign}(dw_U / dp_X) = \mathrm{sign}(R_B - R_A), \qquad \mathrm{sign}(dw_S / dp_X) = -\,\mathrm{sign}(R_B - R_A). \end{aligned}$$

How do the introduction of AI and the exercise of market power affect workers?

National income increases with cheaper AI: applying the envelope theorem to national income, we have $dY/dp_X = -M_X \leq 0$. Further, one wage rises while the other falls when p_X changes; which factor gains depends on the relative intensity of A and B. A change in p_X affects R_B through both substitution across sectors as well as across factors. If B is *unskilled-intensive* ($R_B < R_A$) and p_X increases, then w_S rises and w_U falls; activity tilts from B toward A (which is relatively more skill-intensive).

How does market power interact with income distribution?

Given R_B decreasing in p_X, let the crossing point with R_A (if it exists) be denoted p_X^*. For $p_X < p_X^*$, B is skill-intensive and low AI prices can depress w_U, but increases in p_X then help. For $p_X > p_X^*$, both of these forces reverse.

This is the familiar trade intuition that the losers from ongoing technology adoption in the skill-intensive region may be locally helped by a higher p_X (e.g., a tariff or a markup reversal). However, if unskilled workers are hurt by the introduction of AI, its price must be low enough that Sector B is skill-intensive, so that higher p_X helps the unskilled.

Double vs. single reversals.

Local incidence of AI input prices on the unskilled wage: $\frac{dw_U}{dp_X}$ *vs.* p_X. *Above the zero line, unskilled wages rise with* p_X*; below, they fall. A monotone path yields a single crossing of wages as* p_X *changes (blue), while a non-monotone path yields two crossings (red).*

$R_B(p_X)$ need not be monotone, and for large changes in technology, it may not be. In autarky, sector B may be unskilled-intensive. As the AI price falls, it becomes skill-intensive, and at very low prices it may again appear unskilled-intensive, as AI substitutes for both labor types. Then, the introduction of AI supplied by a firm with market power can deliver a "double-harm": for sufficiently low p_X, the unskilled wage can lie *below* its autarky value *and* be locally *decreasing* in p_X. The local decrease arises because at those very low prices B is again unskilled–intensive relative to A, so a higher p_X reduces B's output and differentially reduces demand for unskilled labor.

A Nontraded Good and Dutch Disease

Now extend the benchmark model to consider a nontraded good A and a traded good B, with Cobb–Douglas preferences. Aggregate real income (or welfare) satisfies

$$d\log W = d\log Y - (1-\alpha)\, d\log p_A,$$

where p_A is the nontraded price. We refer to [eq:RW] as the *real-wage filter*. The mechanism is well known: if the nontraded price rises sufficiently relative to income, both groups' real wages can fall, something which cannot occur when all goods are traded.

We assume that AI only lowers costs in B. The Online Appendix presents comparative statics of p_X on income Y and the traded price $\bar{p}_B$, and then on real income via [eq:RW]. This analysis illustrates the classic logic: as p_X falls, production of the export good B expands, $\bar{p}_B$ rises, and both groups' real wages can fall simultaneously if the CPI channel dominates. This is the textbook Dutch disease result, which refers to the experience of countries that discover a valuable natural resource such as oil (Corden and Neary 1982; Corden 1984). The booming resource sector attracts labor and capital away from other industries, raising national income but also making nontraded goods more expensive. This squeezes households through higher costs of living and erodes competitiveness in other tradables. With AI, the dynamics differ: adoption raises productivity in one sector but pushes, rather than pulls, workers into the rest of the economy. In addition, AI shocks may affect the whole world, making the assumption that the assumption that the country can increase exports at fixed prices a poor fit.

This benchmark also incorporates the logic of Baumol and Bowen (1966): when productivity rises in some sectors but not others, wages equalize across the economy at higher levels (supported by rising aggregate income) and stagnant sectors like education and care see rising labor costs, leading to higher prices, lower welfare, and distributional harm. These forces are not what we believe are most salient for AI. In our main model, technology adoption expands sector B while simultaneously displacing its workers. Those workers are pushed into non-tradables, where their marginal productivity and thus wages fall. If AI is cheap and efficiency gains remain domestic, higher income boosts demand for non-tradables, raising their relative prices and worsening the cost of living. If instead a monopolist extracts large fees, national income may not increase and workers face lower wages without the offsetting demand-driven price surge. However, market power also reduces the extent to which labor cost savings in B pass through: prices in B may fail to fall and variety may shrink.

Main Model

This model builds on the benchmarks to address the issues raised above. We incorporate two new features. First, *both* sectors are non–traded, so that output prices and the expansion of the AI-augmented sector are limited by diminishing marginal utility of consumers. Second, sector B consists of many differentiated varieties under Constant Elasticity of Substitution (CES) demand (elasticity $\sigma > 1$) with constant markup $\mu = \sigma/(\sigma - 1) > 1$ and free entry subject to a per-firm domestic license fee F (rebated lump-sum to households) and a fixed access fee, ϕ, that is collected by the foreign monopolist. This allows us to consider the impact of AI on industry structure and entry, and further opens the door for more realistic pricing strategies by the AI monopolist. The restriction to CES preferences is for expositional simplicity; in the Online Appendix we show that our main results extend to the more general case of Hierarchical Structure of Aggregation (HSA) preferences following Matsuyama (2019).

Studying AI in a fixed-cost framework is natural, since digital technologies involve high up-front investments (training, deployment, access) but low marginal costs of use. A differentiated product and free-entry structure makes it possible to analyze how AI pricing reshapes industry structure (the equilibrium number of firms, variety, and the quality-adjusted price index). Moreover, because AI is a general-purpose technology that enables a wide range of applications, entry and variety are themselves first-order welfare channels.

Primitives and price indices.

Households spend a constant share $\alpha \in (0,1)$ on B and $1 - \alpha$ on A:

$$U = C_A^{1-\alpha} C_B^{\alpha}, \qquad E_B = \alpha Y, \quad E_A = (1-\alpha)Y.$$

Within B, symmetry across the N active firms that the CPI is

$$P = p_A^{1-\alpha} p_B^{\alpha}.$$

Without loss of generality, we normalize $p_A \equiv 1$ so that all prices are relative to the price index in Sector A; although at times we discuss the price effects in Sector A, these should be interpreted as relative prices.

Pricing and free entry in B.

Let m_B be the unit marginal cost and $p_B = \mu\, m_B$ the symmetric price. With per-firm outlays $F + \phi$ and free entry,

$$(p_B - m_B)q = (F + \phi) \quad \Rightarrow \quad p_B\, q = \mu/\mu - 1\,(F + \phi).$$

Because the expenditure in Sector B is $E_B = \alpha Y = N p_B q$, we have

$$N = \mu - 1/\mu\ \alpha\, \frac{Y}{F + \phi}, \qquad P_B = \mu\, m_B \left(\mu - 1/\mu\ \alpha\, \frac{Y}{F + \phi}\right)^{\frac{1}{1-\sigma}}.$$

Thus, holding Y and $(F + \phi)$ fixed, changes in m_B move p_B but do *not* move N directly. In general equilibrium, however, p_X and ϕ move Y and hence N through free entry.

AI technology and marginal cost in B.

A foreign upstream supplier charges a per-firm access fee ϕ and a per-unit usage fee p_X (we start by setting $\phi = 0$ to facilitate comparisons to the benchmarks, and then generalize the model). Adoption in B is discrete and leads to marginal cost $m_B = s_B w_S + p_X$ with no need for unskilled labor ($u_B = 0$), where we say automation is partial if $s_B > 0$ and (informally) full if s_B is significantly lower than the no-AI baseline. For simplicity of exposition, we consider the case where under partial automation s_B is at least as large as the skilled labor used per unit in the benchmark without AI. The case where marginal cost is zero ($s_B = 0$) is an edge case given our assumptions about free entry so we rule it out for simplicity. In both cases, adoption implies $u_B = 0$, so displaced unskilled labor reallocates to A. In contrast to the benchmarks, a marginal expansion of B does not *pull* U from A. The effect of p_X on prices now arises primarily through p_B (via m_B and N), not through p_A (via Q_A).

Sector A and factor markets.

Following the benchmarks, sector A is produced competitively with both skilled and unskilled labor, and labor markets clear:

$$L_S = L_S^A + L_S^B, \qquad L_U = L_U^A + L_U^B.$$

With diminishing returns in A, the crowding of unskilled labor into A lowers w_U, while w_S reflects the allocation of L_S across sectors and the demand for L_S in B. For some results below, we assume that production in Sector A is CES with factor shares (where factors are skilled and unskilled labor) parameterized by β and substitution parameter ρ.

Income and external payments.

Domestic income includes wages and the rebated F but excludes foreign AI payments:

$$Y = w_S L_S + w_U L_U + FN, \qquad \text{Foreign outflow} = \phi N + p_X Q_B.$$

Unlike F, ϕ reduces both entry and *domestic absorption*.

Exogenous AI Prices

Adoption frontier defined.

AI is adopted only if it beats the baseline when rivals adopt. As we discuss in more detail below, this yields a downward-sloping frontier of the set of incentive-compatible (ϕ, p_X) pairs, where the frontier can be written as $p_X \leq p_{X,\max}(\phi)$. Importantly, adoption constraints are more challenging when other firms adopt, since adoption frees up labor and pushes down wages. In this section, we first consider comparative statics on usage and access fees within the frontier, and then return to consider the frontier in more detail when we consider the problem faced by a monopolist AI provider.

Equilibrium.

Let $Z - (\alpha, \sigma, F, s_B;\ L_S, L_U)$ denote the vector of exogenous primitives (preferences, technologies, and endowments) for Model 3. Given (Z, ϕ, p_X), a (competitive) equilibrium consists of prices, quantities, and allocations

$$(p_A,\ P_B,\ p_B,\ m_B,\ N,\ q,\ Q_B;\ w_S, w_U;\ L_S^A, L_S^B, L_U^A, L_U^B;\ Y)^{*r},$$

where the superscript ^{*r} denotes equilibrium values in Model 3 for regime $r \in \{AI, 0\}$ ($r = 0$ being no AI), that jointly satisfy equations [eq:M3-utility-budget]-[eq:M3-income-outflow]. The feasible (ϕ, p_X) must also satisfy an incentive constraint for technology adoption $p_X \leq p_{X,\max}(\phi)$, developed in more detail below. For any equilibrium variable X in the list above, we use the superscript notation $X^{*r}(Z, \phi, p_X)$ to denote its equilibrium value as a function of (Z, ϕ, p_X).

Welfare Ratio

In the Online Appendix, we show that substituting in equilibrium conditions the following expression for the ratio of welfare (real income) across regimes:

$$\boxed{\frac{W^{*AI}}{W^{*0}} = \underbrace{\left(\frac{m_B^{*0}}{m_B^{*AI}}\right)^{\alpha}}_{\text{unit cost / } B \text{ price index}} \underbrace{\left(\frac{F}{F+\phi}\right)^{\frac{\alpha}{\sigma-1}}}_{\text{variety / entry}} \underbrace{\left(\frac{Y^{*AI}}{Y^{*0}}\right)}_{\text{aggregate income}}}$$

That is, the welfare ratio can be decomposed into the product of three ratios. The first is proportional to the ratio of marginal costs across regimes, which affects the welfare ratio through the price index; the second is the ratio of fixed costs, which affects the ratio of the number of firms and thus variety; the third is proportional to the ratio in aggregate income. (Recall that p_A is normalized to 1 so that income is considered relative to p_A). Each of these terms depends on the fees (ϕ, p_X). At low/zero prices, unit costs fall, the variety/entry effect is equal to one, and equilibrium income rises. On the other hand, at high access or usage fees (but low enough to induce adoption), the unit cost/price index effect is dampened, the variety/entry effect reduces the ratio, and income is reduced.

The same decomposition can also be applied separately for skilled and unskilled workers, where the first two terms are the same across worker groups. However, wage shifts due to the adoption of AI drive differences in the income ratio changes. Displaced workers may have lower labor income, so that unskilled workers may be harmed even by cheap AI, depending on parameter values, similar to the analysis of the benchmark models.

Linear pricing (usage–only, $\phi = 0$).

Under linear pricing the foreign supplier sets the per–unit usage fee p_X with no access fee ($\phi = 0$).

We model the introduction of the technology as a shift from autarky ($p_X = \infty$) to finite p_X. Two new forces arise relative to the benchmarks. First, a *variety channel*: lower p_X reduces unit costs in B, raises output and income, and—through free entry—supports more firms and more varieties. More varieties lower the quality–adjusted price index P_B, a channel absent when B was traded at a fixed world price. Second, a *displacement channel*: once AI is adopted, all L_U is pushed into sector A. With Cobb–Douglas in A, this crowding reduces w_U through diminishing returns, an additional effect beyond the benchmark.

In equilibrium, a lower p_X reduces unit costs in B and raises variety. Under partial automation, the additional production in B pulls L_S out of A, which pushes p_A upward. Under mild conditions, welfare rises relative to autarky, as the unit–cost and variety gains outweigh the rise in p_A. But unskilled workers' real wage typically falls, since the unskilled nominal wage is depressed by crowding in A while the CPI rises with p_A. Skilled workers may gain under partial automation (more demand for L_S in B), but under full automation, they too may lose as almost all labor is absorbed into A while gains leak abroad through p_X.

A small increase $dp_X > 0$ when the AI adoption constraint is slack. Suppose the adoption constraint is slack. Then, raising p_X increases unit costs in B and contracts its output. Because B has CES demand and free–entry, this also reduces the number of firms, raising P_B via the variety channel. At the same time, fewer firms in B release some L_S back to A, lowering p_A.

Unlike the benchmarks, these local changes do not induce substitution back to L_U in B: adoption is discrete, so a marginally higher p_X worsens P_B without undoing displacement. When the expenditure share α is moderate, the worsening of P_B and the fall in income dominate any relief in p_A, so that the CPI rises and welfare falls. In this case both w_S/P and w_U/P typically decline. In the Online Appendix, we analyze these comparative statics in the case of a CES production function in Sector A. The outcomes are most stark when the two types of labor are highly substitutable in sector A: the relief in p_A vanishes, and both real wages fall unconditionally. Conversely, when the production function in A is close to Leontief, unskilled labor can be partially cushioned or even gain from higher ϕ (Corollary [cor:rho-access]), so the negative incidence is less certain. Thus, as long as α is not too small and A is not extremely unskilled-intensive, the local and global incidence can align negatively for at least one worker group, reproducing the "double harm" result highlighted in Models 1 and 2.

Two–part tariffs (ϕ, p_X).

Allowing the foreign AI supplier to set both a per–unit price p_X and an access fee ϕ adds a second lever. A higher ϕ directly reduces entry in B, raising P_B, and, unlike F, ϕ's proceeds leak abroad, lowering domestic income.

Relative to the benchmark models without foreign ownership, the CES environment introduces two additional channels when the AI supplier raises p_X. First, higher unit costs and lower variety both raise P_B, strengthening the CPI channel. Second, because profits include foreign revenues, a new income–leakage effect arises: as p_X rises, domestic license rebates FN shrink, depressing Y. Together these channels imply the price index rises and real wages fall. Moreover, the loss of domestic license rebates increases inequality across workers: all lose in real terms, but the erosion of FN magnifies relative differences between skilled and unskilled labor.

Overall, the combination of discrete displacement, endogenous variety, and two–part tariffs makes it far easier for market power in AI supply to depress welfare, and it is more likely that the introduction of AI harms unskilled workers, and that local exercise of market power exacerbates the harm.

AI Monopolist Choice of Usage and Access Fees

Sector B Firm Profits.

We can also examine the impact of access and usage fees on Sector B production and gross profits. Define the gross firm–side aggregate operating surplus before fees as follows (where we substitute in equilibrium conditions):

$$\mathcal{E}^{\text{gross}}(Z,\phi,p_X) \;\equiv\; (\mu-1)\, m_B^{*AI}(Z,\phi,p_X)\, Q_B^{*AI}(Z,\phi,p_X) = \frac{(\mu-1)\alpha}{\mu}\, Y^{*AI}(Z,\phi,p_X).$$

Note that AI access and usage fees impact this *only* through Y^{*AI}. If we think of

$$\mathcal{E}^{\text{gross}}(Z,\phi,p_X) - N^{*AI}F = \frac{(\mu-1)\alpha}{\mu}\, Y^{*AI}(Z,\phi,p_X)\left(\frac{\phi}{F+\phi}\right)$$

as the "size of the pie" to be extracted by the monopolist through access and usage fees, we can see that both ϕ and p_X affect the size of the pie through their effect on income. This contrasts with the typical nonlinear pricing problem from industrial organization, where the number of firms is fixed and there are no general equilibrium effects, so that access fees do not distort production while usage fees do. In the latter case, it is optimal for a monopolist to keep usage fees as low as possible and extract surplus using access fees; in contrast, in our model both fees increase the share of profits that go to the AI monopolist.

The slope of income with respect to ϕ is

$$\frac{\partial \mathcal{E}^{\text{gross}}}{\partial \phi}(Z,\phi,p_X) = \frac{(\mu-1)\alpha}{\mu}\,\frac{dY^{*AI}}{d\phi}(Z,\phi,p_X) \;<\; 0,$$

where $dY^{*AI}/d\phi$ incorporates a *direct* income effect, where ϕ affects the share of consumer expenditure retained by firms, and an *indirect* effect via the induced change in Q_B^{*AI}. In contrast, p_X has only an indirect effect on $\mathcal{E}^{\text{gross}}$, through wages and output that affect income.

Adoption frontier in equilibrium.

To analyze strategic pricing by the monopolist, we develop the adoption frontier $p_X \leq p_{X,\max}(\phi)$ in more detail. When rivals adopt, wages (w_S^{*AI}, w_U^{*AI}) fall for the displaced factor, so the baseline alternative improves; sustaining adoption therefore requires *lower* p_X or *lower* ϕ.

Consider one firm that *deviates* to the baseline (non-adopting) technology while all rivals keep adopting and charging $p_B^{*AI} = \mu\, m_B^{*AI}$. Let the deviator set the usual markup price $p_{\text{dev}} = \mu\, m_{\text{dev}}$, where its baseline marginal cost is evaluated at the AI equilibrium wages:

$$m_{\text{dev}}(Z,\phi,p_X) \;\equiv\; s_B\, w_S^{*AI}(Z,\phi,p_X) \;+\; u_B\, w_U^{*AI}(Z,\phi,p_X).$$

Under CES demand and substituting in equilibrium conditions, the deviator's quantity at (Z,ϕ,p_X) is

$$q_{\text{dev}}(Z,\phi,p_X) = \alpha\, Y^{*AI}\, (\mu\, m_{\text{dev}})^{-\sigma}\, (P_B^{*AI})^{\sigma-1} = \frac{\alpha}{\mu}\frac{Y^{*AI}}{m_{\text{dev}}}\left(\frac{m_B^{*AI}}{m_{\text{dev}}}\right)^{\sigma-1}\frac{1}{N^{*AI}}.$$

The deviator does not pay the AI access fee (it does not adopt), so its profit is

$$\pi_{\text{dev}}(Z,\phi,p_X) = (\mu-1)\, m_{\text{dev}}\, q_{\text{dev}} - F = \frac{(\mu-1)\alpha}{\mu}\, Y^{*AI}\left(\frac{m_B^{*AI}}{m_{\text{dev}}}\right)^{\sigma-1}\frac{1}{N^{*AI}} - F.$$

Using [eq:M3-N-and-PB], the no-deviation condition $\pi_{\text{dev}} \leq 0$ is equivalent to

$$\left(\frac{m_B^{*AI}}{m_{\text{dev}}}\right)^{\sigma-1} \leq F/F+\phi \;\Leftrightarrow\; p_X \leq m_{\text{dev}}(F/F+\phi)^{\frac{1}{\sigma-1}} - s_B w_S^{*AI}.$$

Usage fee.

Consider linear pricing with $\phi = 0$, so the monopolist's revenue is

$$\Pi_{\text{lin}}(p_X) \;=\; p_X\, Q_B \;=\; \frac{\alpha}{\mu}\frac{Y}{m_B}\, p_X,$$

where $m_B = s_B w_S + p_X$ and $Q_B = (\alpha/\mu)(Y/m_B)$ in equilibrium. Differentiating shows that profits rise with p_X so long as the negative impact of p_X on income Y is not too strong relative to the positive cost-share transfer from skilled labor to the monopolist. In this case the profit function is increasing up to the adoption cap, so the optimal usage fee is set at the boundary.

Intuitively, when sector B still requires some skilled labor ($s_B > 0$), a higher p_X both raises m_B and shifts part of the cost burden away from domestic wages, increasing the monopolist's margin. The opposing force is the contraction of income Y, which lowers overall expenditure on B. Profits rise with p_X provided this income contraction is not too severe. This condition is more likely to hold when the expenditure share α on B is moderate, when skilled labor's cost share in m_B is sizable, and when demand for B is relatively elastic (so markups μ are modest). It is also easier to satisfy when technology in sector A allows factors to substitute smoothly: in that case the fall in skilled wages is cushioned, the rise in unskilled wages is limited, and the overall income decline is modest. By contrast, when technology in A is close to fixed-proportions, the rise in p_X depresses Y more strongly, making it harder for the cost-share transfer to dominate. Note, however, that at intermediate levels of ρ (the parameter governing factor substitution in Sector A), outcomes can be nonmonotone in ρ.

Access fee.

Now consider the AI monopolist's choice of both usage and access fees,

$$\Pi^{\text{acc}}(Z, \phi \mid p_X) \;\equiv\; \underbrace{\phi\, N^{*AI}(Z, \phi, p_X)}_{\text{access revenue}} \;+\; \underbrace{p_X\, Q_B^{*AI}(Z, \phi, p_X)}_{\text{usage revenue}}.$$

Using the CES/free–entry identities, differentiation gives

$$\frac{\partial \Pi^{\text{acc}}}{\partial \phi} = N^{*AI}\left[\frac{F}{F+\phi} \;-\; \frac{\phi}{Y^{*AI}}\frac{dY^{*AI}}{d\phi} \;-\; \frac{p_X}{(\mu-1)\, m_B^{*AI}} \cdot \frac{F+\phi}{Q_B^{*AI}} \cdot \frac{dQ_B^{*AI}}{d\phi}\right].$$

The first term is strictly positive. The profit slope in ϕ reflects a direct positive channel and indirect negative channels through Y and Q_B. When adoption is slack, any interior optimum $\phi^\star(Z \mid p_X)$ requires these effects to balance exactly. If the adoption frontier is binding, the optimum lies on the boundary at the highest ϕ consistent with adoption.

The strength of the indirect terms depends on how easily factors can reallocate in A. When technology in A allows smooth substitution between skilled and unskilled labor, the contraction in Y from a higher ϕ is relatively muted. In this case, the direct access-revenue channel dominates, so interior optima are less likely and the monopolist tends to push ϕ to the frontier. By contrast, when A is closer to fixed proportions, the fall in income is sharper, the negative terms dominate sooner, and an interior optimum in ϕ is more plausible. At the extreme, with highly substitutable A, the indirect contraction is small enough that unskilled wages can still fall with ϕ, while with fixed-proportions A, unskilled wages may rise in nominal terms even as welfare declines. Thus, with smooth substitution in A the monopolist relies more on the per–unit fee, while with rigid A technology the fixed access fee is relatively more attractive to the monopolist. At intermediate levels of ρ, outcomes can be non-monotone in ρ.

Summary of results

Monopoly power in AI depresses welfare and can harm both skilled and unskilled workers. Relative to standard benchmarks, several mechanisms are distinctive here:

1. *No unskilled pull in B.* Adoption is discrete, so marginal fee changes do not restore unskilled demand in B; displaced labor must be absorbed by A.
2. *Two CPI channels.* The per–unit usage fee raises unit costs in B, while the fixed access fee reduces the number of active firms and thus variety. Both channels raise the sector-B price index and contribute to CPI inflation.
3. *Income leakage.* Both fees transfer income abroad, leading to declines in national income and a more concentrated domestic industry with less variety.
4. *Distributional impacts.* Because all workers face the common CPI, real wages for both skilled and unskilled typically fall when AI fees rise. Skilled wages decline in both nominal and real terms. Unskilled nominal wages can rise with higher p_X, but real wages usually fall once the CPI effect is taken into account. Access fees add an additional distributional margin by eroding domestic license rebates, which magnifies inequality between workers under our assumption that all workers share in them.
5. *Role of technology in A.* The qualitative incidence depends on how easily factors substitute in A. When A is highly substitutable, income contractions from either AI fee are muted but CPI pressures dominate, so both groups lose in real terms unconditionally. When A consumes labor types in close to fixed proportions, unskilled labor may gain nominally from higher access fees, so the incidence is less stark.

In all cases, the monopolist allocates extraction across fixed and variable fees, adjusting one upward if the other is capped. Regulating only one instrument in isolation therefore does not guarantee an improvement in national outcomes. Monopoly power in a general-purpose technology affects not only wages and aggregate income but also the range of products and the quality-adjusted consumption basket available to consumers.

A Policy-Relevant Research Agenda

The models developed above are deliberately stylized, but we argue that they help identify and prioritize open questions for research and considerations for policy-makers. First, in the models, broad welfare gains require that *AI prices fall with real cost savings* (either through competition enforcement or regulation of p_X and ϕ) *and* that displaced labor be *absorbed productively* in sectors less affected by AI. Second, we show that policies that only cap one instrument risk rent-shifting into the other; disciplining both levers is necessary when the upstream supplier prices on the adoption frontier.

From a modeling perspective, the simple structure of CES/HSA demand and a monolithic AI industry could both be generalized, yielding more nuanced insights. A long tradition in industrial organization (see, e.g. (Lee et al. 2021) for a recent survey) and a recent literature in trade (e.g. (Grossman et al. 2024)) microfound markups in supply chains, and the latter explores general equilibrium implications. From a country's perspective, a key factor is where AI value added accrues. If domestic firms earn the profits and those profits recirculate broadly in the economy, they raise national income and help sustain demand for non-tradables. Future research might consider what policies enable a country to participate in the high-markup parts of the AI stack. Competition policy and industrial policy might be able to improve a country's bargaining leverage with outside suppliers. For example, a country may be able to control access to data or financial systems. It might consider whether open models (such as Meta's Llama) should be regulated for national security reasons (though the arrival of DeepSeek demonstrated that it may be difficult to keep countries from fast-following others' innovations). Outcomes may depend on where a country's local assets (e.g., data, finance, compute, talent) create favorable outside options.

What policies increase competition? In some cases, multiple global providers may exist, but not all make investments in customizing and distributing to small countries, reducing local competition and leaving a role for industrial policy. The AI stack spans many layers, including chips, training, models, and applications. A single bottleneck can lead to high downstream prices. Where bottlenecks are likely, how pricing at each layer translates into end-user prices and pricing structure, and which policy tools best counteract market power are all important research topics. For example, open source models have a role to play in constraining market power. The presence of open models, competitive entrants, or fast followers might discipline incumbents and reduce the scope for sustained market power.

Another risk is capability loss: if adoption displaces domestic production that is costly to rebuild (learning-by-doing, organizational capital, supply-chain agglomeration), dependence on external suppliers increases. A country's outside option, and therefore ability to bargain over and regulate AI import prices, may deteriorate. That raises more industrial-policy questions: how should short-run efficiency gains be weighed against longer-run resilience and expertise? Closely related is the risk that a dominant AI provider falls under control of a hostile trading partner, so that objectives other than profit—e.g., degrading local capabilities—become salient.

For workers, the key question is whether displaced labor can be reallocated productively. Policy levers include education and training, procurement of nontraded services (teaching, nursing, care), and targeted subsidies that raise productivity where displaced workers are absorbed. AI assistants may ease both transitions and on-the-job productivity. The political economy around these large fiscal and operational decisions for government will have a big impact on outcomes such as who pays, how efficient redistribution is, and whether AI owners shape policy to avoid bearing social costs. Concentrated ownership, even by domestic firms, may increase their political power and make taxation more difficult.

In standard models, if AI-enabled industries are exportable, higher productivity and increased output mitigate the decreases in per-unit labor demand. However, this result depends on the ability to increase exports at similar prices, when in practice world-wide adoption of AI may compress prices and limit export-led adjustment. Some traded services (e.g., call-center outsourcing) may shrink materially. The implications for developing countries' competitiveness and the global distribution of production are critical to understand.

The interaction between AI and innovation itself also opens a rich research agenda. On the one hand, AI can erode market power in existing downstream industries by lowering costs and enabling entry. On the other hand, our analysis shows that if fixed costs rise, industry structure may become more concentrated, reducing the variety available to end consumers and slowing innovation. From the perspective of AI innovation, if monopoly rents are the primary reward for investing in frontier AI, restricting those rents too aggressively might dull incentives for innovation. Understanding how different forms of pricing, competition, and regulation shape innovation incentives in AI, and how this interacts with broader patterns of technological progress, remains a pressing challenge.

Conclusions

This paper argues that the payoff from preserving competition in AI has been underappreciated in macroeconomic discussions. Growth models and popular narratives often assume that AI will deliver cheap goods and services in abundance. Our observation is that this outcome is unlikely in the absence of competition, particularly if the profits from the technology accrue outside a country or are not shared throughout the economy. Instead, we show how monopoly in AI allows the provider to extract rents and prevent prices from falling to match declines in labor. A profit-maximizing downstream firm may be just indifferent about adopting AI, but the decision redistributes rents outside the country and away from displaced workers within the country. Because AI is a general-purpose upstream technology, monopoly harms include increased concentration within downstream sectors but extend well beyond directly impacted sectors, justifying a general equilibrium framework. Our models demonstrate that the welfare impact of AI depends critically on market structure; they further highlight the important role of worker transitions and their unique value and productivity in alternative sectors.

Aghion, Philippe, Benjamin F. Jones, and Charles I. Jones. 2019. "Artificial Intelligence and Economic Growth." In *The Economics of Artificial Intelligence: An Agenda*, edited by Ajay Agrawal, Joshua Gans, and Avi Goldfarb. University of Chicago Press.

Autor, David H., Frank Levy, and Richard J. Murnane. 2003. "The Skill Content of Recent Technological Change: An Empirical Exploration." *The Quarterly Journal of Economics* 118 (4): 1279–333. https://doi.org/10.1162/003355303322552801.

Baqaee, David, and Emmanuel Farhi. 2019. "The Macroeconomic Impact of Microeconomic Shocks: Beyond Hulten's Theorem." *Econometrica* 87 (4): 1155–203. https://doi.org/10.3982/ECTA15202.

Baqaee, David, and Emmanuel Farhi. 2020. "Productivity and Misallocation in General Equilibrium." *The Quarterly Journal of Economics* 135 (1): 105–63. https://doi.org/10.1093/qje/qjz030.

Baumol, William J., and William G. Bowen. 1966. *Performing Arts: The Economic Dilemma*. Twentieth Century Fund.

Competition and Markets Authority. 2023. *AI Foundation Models Technical Update Report*. https://assets.publishing.service.gov.uk/media/65081d3aa41cc300145612c0/Full_report_.pdf.

Corden, W. Max. 1984. *Booming Sector and Dutch Disease Economics: Survey and Consolidation*. Oxford Economic Papers on Energy No. 3. Vol. 36. Oxford Institute for Energy Studies.

Corden, W. Max, and J. Peter Neary. 1982. "Booming Sector and de-Industrialisation in a Small Open Economy." *The Economic Journal* 92 (368): 825–48. https://doi.org/10.2307/2232670.

Delli Gatti, Domenico, Mauro Gallegati, Bruce C Greenwald, Alberto Russo, and Joseph E Stiglitz. 2012. "Sectoral Imbalances and Long-Run Crises." In *The Global Macro Economy and Finance*. Springer.

Domar, Evsey D. 1961. "On the Measurement of Technological Change." *The Economic Journal* 71 (284): 709–29. https://doi.org/10.2307/2228246.

Gatti, Domenico Delli, Mauro Gallegati, Bruce C Greenwald, Alberto Russo, and Joseph E Stiglitz. 2012. "Mobility Constraints, Productivity Trends, and Extended Crises." *Journal of Economic Behavior & Organization* 83 (3): 375–93.

Grossman, Gene M., Elhanan Helpman, and Alejandro Sabal. 2024. "Optimal Resilience in Multi-Tier Supply Chains." *Quarterly Journal of Economics* 139 (4): 2377–425. https://doi.org/10.1093/qje/qjae024.

Hulten, Charles R. 1978. "Growth Accounting with Intermediate Inputs." *The Review of Economic Studies* 45 (3): 511–18. https://doi.org/10.2307/2297332.

Korinek, Anton, and Joseph E Stiglitz. 2018. "Artificial Intelligence and Its Implications for Income Distribution and Unemployment." In *The Economics of Artificial Intelligence: An Agenda*. University of Chicago Press.

Korinek, Anton, and Joseph E Stiglitz. 2021. *Artificial Intelligence, Globalization, and Strategies for Economic Development*. National Bureau of Economic Research.

Lee, Robin S., Michael D. Whinston, and Ali Yurukoglu. 2021. “Structural Empirical Analysis of Contracting in Vertical Markets.” In *Handbook of Industrial Organization, Volume 4*, edited by Kate Ho, Ali Hortaçsu, and Alessandro Lizzeri, vol. 4. Elsevier. https://doi.org/10.1016/bs.hesind.2021.11.009.

Matsuyama, Nobuhiro. 2019. “Engel’s Law in the Global Economy: Demand-Induced Patterns of Structural Change, Innovation, and Trade.” *Econometrica* 87 (2): 497–528.

Nordhaus, William D. 2021. “Are We Approaching an Economic Singularity? Information Technology and the Future of Economic Growth.” *American Economic Journal: Macroeconomics* 13 (1): 299–332. https://doi.org/10.1257/mac.20170105.

Portuguese Competition Authority. 2023. *Competition and Generative Artificial Intelligence*.

Stiglitz, Joseph E. 1976. “Monopoly and the Rate of Extraction of Exhaustible Resources.” *The American Economic Review* 66 (4): 655–61.