Generative AI as an Inequality Amplifier: Language Compatibility, Leapfrogging and Productivity Growth

Yang Chen Anxu Wang

June 2026

Abstract

We provide the first causal evidence that generative AI amplifies global inequality through linguistic compatibility with training data. Exploiting the November 2022 release of ChatGPT as a natural experiment across 169 economies, we find that a one-standard-deviation increase in Language Intensity—a country's linguistic compatibility with AI training data—raises productivity growth by 0.94 percentage points. This produces "linguistic leapfrogging": low-income but linguistically compatible economies—Ghana, Nigeria—match the gains of wealthy Anglophone nations, while high-income but incompatible economies like Japan see none. We conduct a controlled experiment to provide micro-foundations: across languages, a generative AI tool varies widely in both the quality and the cost of its business decisions, leaving substantial value unrealized where it performs worst. A structural model of AI-augmented technology adoption, calibrated to our empirical estimate, projects these patterns forward and shows that multilingual AI development would cut the inequality cost of AI-driven growth by 59 percent. Our results challenge standard views of technology diffusion and reveal that AI introduces a new axis of global stratification along linguistic lines.

Keywords: Artificial Intelligence, Global Inequality, Technology Diffusion, Language Compatibility, Productivity Growth

JEL Codes: O33, O15, O19, O31

Yang Chen: School of Innovation and Development, Central University of Finance and Economics (2023212533@email.cufe.edu.cn). Anxu Wang: National School of Development, Peking University (axwang2023@nsd.pku.edu.cn).

For valuable comments and suggestions we thank Yang Yao, Lixing Li, Tianyang Xi, Xuan Wang, Jiaqi Zhao and seminar participants at the National School of Development, Peking University. The authors declare that they have no relevant financial relationships or other potential conflicts of interest. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

1 Introduction

The diffusion of new technologies has reshaped the global income distribution at every major turning point in economic history. From electricity to telecommunications to the internet, each general-purpose technology has redrawn the map of who leads and who lags (Comin and Hobijn, 2010; Comin and Mestieri, 2018). A longstanding debate asks whether new technologies narrow or widen global inequality: the Gerschenkron (1962) tradition emphasizes that latecomers can leapfrog by adopting frontier technologies directly, while a large empirical literature documents that wealthy nations systematically adopt first and thereby pull ahead. Acemoglu et al. (2026) sharpen this tension by showing that when frontier technologies are "inappropriate" for developing economies' conditions, adoption slows and income gaps widen—technology appropriateness, not just access, determines who benefits. Artificial intelligence is the current episode in this history. But is the pattern the same?

Recent evidence documents substantial productivity gains from generative AI adoption within firms and tasks (Brynjolfsson et al., 2025; Noy and Zhang, 2023; Dell'Acqua et al., 2023). ChatGPT reached 100 million users in two months—faster than any technology in history—and is rapidly spreading across economies at all income levels, making AI a phenomenon that demands macroeconomic and cross-country analysis. Yet whether these micro-level gains aggregate into measurable economy-wide productivity growth remains contested. Existing estimates range from 0.07 to 1.3 percentage points annually (Acemoglu, 2025; Aghion and Bunel, 2024; Filippucci et al., 2025), all derived from calibration rather than realized data. The distributional question is in even worse shape. Policy reports warn of a new "Great Divergence" (Council of Economic Advisers, 2026) and theoretical work predicts AI may widen cross-country income gaps (Acemoglu et al., 2026), but causal evidence on which economies actually gain and which fall behind is altogether absent.

This paper fills that gap by exploiting two sources of variation. First, ChatGPT's November 2022 release provides a sharp temporal shock—one so consequential that disruption at this scale now has a name: "ChatGPT moment." Second, we leverage cross-economy varia-

tion in linguistic compatibility with AI training data—a natural dimension given how large language models work. LLMs generate text by predicting the next most probable token based on patterns in training corpora. Since Common Crawl, the primary corpus for GPT models, is 44.3 percent English and less than 5 percent most other languages, models naturally perform better in well-represented languages (Ahuja et al., 2023; Bang et al., 2023).

Language compatibility is a particularly well-suited treatment intensity for studying AI's productivity effects, for both conceptual and econometric reasons. Conceptually, this is not merely a translation problem. Philosophers from Heidegger to Wittgenstein have long recognized that language constitutes thought itself rather than serving as a neutral medium for pre-formed ideas; AI tools aligned with one's mother tongue thus reduce cognitive friction in ways translation cannot replicate. The policy relevance of this asymmetry is already visible: in January 2026, NVIDIA and AI Singapore jointly released Nemotron-Personas-Singapore, a synthetic dataset explicitly designed to counter the dominance of English-centric training data. More importantly, linguistic composition is shaped by centuries of demographic and colonial history, making it pre-determined with respect to AI's emergence. AI adoption or usage measures, by contrast, are themselves consequences of the technology and cannot support causal identification. Language thus offers a rare opportunity to estimate AI's productivity effects using a treatment intensity that is both economically meaningful and econometrically exogenous.

In this context, we construct a "Language Intensity" measure capturing each economy's linguistic compatibility with AI training data. Examining productivity growth across 169 economies from 2015 to 2024, we find that a one-standard-deviation increase in Language Intensity raises post-ChatGPT productivity growth by 0.94 percentage points, a 58 percent acceleration over baseline. This provides the first causal evidence that generative AI generates measurable economy-wide productivity growth, validating micro-level task studies at aggregate scale. More importantly, this growth is highly unequal: it concentrates entirely among economies whose populations speak languages well-represented in AI training data.

Economies with well-represented languages experience significant acceleration, while linguistically distant economies see no gains, even when wealthy and institutionally strong. AI thus operates as an "inequality amplifier" through linguistic compatibility.

The distributional pattern is unexpected. We categorize economies by income and language compatibility, revealing unexpected winners and losers. Low-income economies with high linguistic compatibility—Ghana, Nigeria, Kenya—experience productivity acceleration comparable to wealthy Anglophone nations like the United States and United Kingdom. Meanwhile, high-income non-English economies—Japan, South Korea, Germany—show no significant gains. This "linguistic leapfrogging" challenges conventional models of technology diffusion, which predict that wealthy economies adopt new technologies first and thereby pull ahead. Here, linguistic advantage appears to trump economic resources.

Several robustness checks strengthen the causal interpretation. Event study estimates show no differential trends in the years before ChatGPT's release, followed by an immediate divergence in 2022 that persists through 2024. We exploit colonial history to address institutional confounds. Restricting analysis to former colonies—where institutional and cultural legacies are broadly shared—linguistic compatibility continues to predict productivity gains. Moreover, comparing across colonizer identity, effects concentrate among English-speaking former colonies while French and Spanish ones show no significant gains, mirroring AI training data composition rather than institutional legacies. We also control for AI adoption using Google Trends data, confirming that our results reflect differential effectiveness conditional on exposure, not merely differences in awareness or uptake. Further analysis reveals that language compatibility is necessary but not sufficient: the productivity effect is concentrated in economies with high internet penetration, pointing to a "double divide" where both linguistic compatibility and digital infrastructure are required.

We next conduct a controlled experiment to provide micro-foundations for our cross-country estimates. The design extends a line of cross-language AI research we have pursued in computational social science (Zhang et al., 2025)—the intellectual starting point of this

paper. We put leading AI models to work on canonical management problems—linear programming (product mix), decision analysis under uncertainty (value of information), and market-entry strategy—posed in languages from English to Yoruba. We score not linguistic accuracy but the economic value of each decision and the cost of producing it. The protocol meets the standards of multilingual AI-simulation research: stateless, temperature-zero queries on identical instances across languages, benchmarked against exactly computed optima. Both the quality and the cost of the models' decisions worsen as the language grows less represented in training data, with economically large losses at the low-resource end. This task-level evidence corroborates our framework's necessary-condition logic: the compatibility that predicts cross-country growth also governs the value AI delivers on a single business decision.

To move beyond what reduced-form estimates can tell us, we develop a structural model extending the technology adoption framework of Acemoglu et al. (2026). We decompose the technology adoption rate into a conventional channel and an AI-augmented channel, where the latter is gated by linguistic compatibility λ_i. The model delivers three propositions: AI productivity gains depend only on linguistic compatibility and are independent of traditional development factors (Proposition 1); language is a necessary condition that wealth and institutions cannot substitute for (Proposition 2); and there exists a threshold of AI importance beyond which poor English-speaking economies overtake wealthy non-English ones (Proposition 3). These predictions map directly onto our empirical findings and provide a structural interpretation of the reduced-form results.

Generative AI is advancing at an unprecedented pace—the capabilities of frontier models have transformed dramatically in the three years since ChatGPT's release—and the inequality patterns we document are only the beginning. Calibrating the model to our empirical estimate (with a single free parameter μ), we conduct counterfactual analyses that project these patterns forward. The quantitative analysis reveals that 84 percent of the world's workers gain less than 4.3 percent from AI at current levels, while 5 percent of the labor

force captures gains exceeding 30 percent. The English-language bias in training data costs an estimated $3.6 trillion in forfeited annual output. Multilingual AI development would cut the inequality cost of AI-driven growth by 59 percent—a Pareto improvement that extends benefits without reducing them for any country. Sovereign AI development, as exemplified by China's DeepSeek, offers a complementary path: economies can expand their own linguistic compatibility through domestic AI investment, producing a bimodal rather than unipolar distribution of AI-driven gains.

We make four contributions. First, we bridge micro-level evidence on AI's productivity impact and its macroeconomic aggregates. Brynjolfsson et al. (2025) find ChatGPT raises customer service worker productivity by 14 percent; Noy and Zhang (2023) show 40 percent time reductions on writing tasks; Dell'Acqua et al. (2023) document gains for BCG consultants; Bick et al. (2026) document rapid AI adoption across the workforce. These studies establish AI's potential at the task level but focus on specific contexts within developed economies, leaving aggregate effects unclear. A parallel literature attempts to translate these micro-level gains into aggregate projections through calibration, with estimates ranging from 0.07 percentage points annually (Acemoglu, 2025) to 0.4–1.3 percentage points for high-exposure economies (Aghion and Bunel, 2024; Filippucci et al., 2025)—an order-of-magnitude disagreement that calibration alone cannot resolve. We provide the first cross-country causal estimate of generative AI's aggregate productivity impact using realized data: a 0.94 percentage point annual acceleration in linguistically compatible economies. This estimate falls within the optimistic range of the calibration literature and substantially exceeds the most conservative projections. A growing literature also examines AI's distributional effects, including Cruces et al. (2026) who find AI narrows productivity gaps across education levels; we contribute a new dimension—linguistic compatibility with training data—along which AI amplifies inequality.

Second, we challenge canonical models of technology diffusion. Comin and Hobijn (2010) document that wealthy nations systematically adopt new technologies earlier, from tele-

graphs to the internet, driving income divergence. Comin and Mestieri (2018) show technology diffusion lags explain much of cross-country income differences. Recent work extends this logic to AI, projecting that gains will concentrate in high-exposure advanced economies while low- and lower-middle-income economies face headwinds from weak digital infrastructure and limited knowledge-intensive services (Filippucci et al., 2025; Chaar et al., 2025). Our findings reveal AI breaks this pattern. Poor English-speaking economies do not lag wealthy Anglophone nations in AI-driven productivity gains; they match them. Conversely, wealthy non-English economies—which led in adopting previous technologies—show no gains. This "linguistic leapfrogging" introduces a novel mechanism: linguistic compatibility enables poor economies to overtake rich ones when new technologies exhibit training data bias. It also identifies a new source of AI-era comparative advantage: where Goldfarb and Trefler (2019) show that AI reshapes international trade through scale and knowledge diffusion, we find that linguistic compatibility with training data redraws the map of who gains. Our theoretical model formalizes this phenomenon, providing an explicit threshold condition under which leapfrogging occurs (Proposition 3).

Third, we bring direct experimental evidence to a mechanism that cross-country data can only probe indirectly. A literature in computer science documents that large language models perform worse in low-resource languages (Lai et al., 2023; Ahuja et al., 2023; Bang et al., 2023), but it measures benchmark accuracy rather than economic value. Continuing the multilingual evaluation framework of Zhang et al. (2025), we run a controlled experiment in which a generative-AI tool makes common business decisions across languages, scored by the economic value of its outputs against a computable optimum—and find that the same linguistic gradient driving our aggregate estimates governs the value of an individual decision. This adds a cross-language dimension to the economics of AI as a decision technology: AI capability is "jagged" not only across tasks (Gans, 2026) but across languages, and degrades most on the judgment-intensive decisions least amenable to machine prediction (Agrawal et al., 2019).

Fourth, we extend the literature on inappropriate technology and AI's distributional consequences. Acemoglu et al. (2026) warn that AI, optimized for advanced economies, may prove inappropriate for developing economies. Moscona and Sastry (2025) show agricultural technologies fail to transfer when pathogen environments differ. We confirm the warning empirically while identifying a more precise mechanism: the relevant dimension of mismatch is not factor endowments but linguistic compatibility with training data—a finding that also implicates wealthy non-English economies as unexpected losers. Our model shows that this dimension of inappropriateness is qualitatively stronger than standard mismatch: language operates as a near-necessary condition (Proposition 2), not merely a complement that reduces gains at the margin. The quantitative analysis translates this into concrete numbers: the annual global cost of linguistic mismatch is $3.6 trillion, and multilingual AI development represents a Pareto improvement that cuts the inequality cost of growth by 59 percent.

Taken together, our findings reveal that AI introduces a new axis of global inequality—one defined not by income or institutions, but by linguistic compatibility with training data. This axis cuts across the traditional rich–poor divide, producing a pattern of linguistic leapfrogging in which poor English-speaking economies match the gains of wealthy Anglophone nations while wealthy non-English economies fall behind. Yet this outcome is not inevitable. Training data composition is a design choice, and our counterfactual analysis shows that broadening the linguistic reach of AI development would extend its productivity gains without diminishing them anywhere.

The paper proceeds as follows. Section 2 develops the theoretical framework. Section 3 describes our data and the construction of Language Intensity. Section 4 outlines the empirical strategy. Section 5 presents baseline results. Section 6 provides robustness checks. Section 7 documents linguistic leapfrogging through heterogeneity analysis. Section 8 provides direct experimental evidence of the mechanism at the task level. Section 9 calibrates the model and conducts counterfactual experiments. Section 10 discusses theoretical and policy implications. Section 11 concludes.

2 Theoretical Framework

This section develops a theoretical framework that motivates and disciplines the empirical analysis. We begin with three micro-level arguments for why linguistic compatibility with AI training data should affect labor productivity (Section 2.1). We then formalize these arguments within a tractable model of cross-country technology adoption, building on the framework of Acemoglu et al. (2026) (Section 2.2–2.6). The model delivers three testable predictions that map directly onto our empirical strategy and results.

2.1 Why Language Matters for AI

Three channels link linguistic compatibility to AI-driven productivity.

Training data composition and model performance. Large language models predict tokens based on patterns learned from training corpora. The dominant corpus for GPT models, Common Crawl, is approximately 44.3 percent English, 5.8 percent Chinese, 4.3 percent Spanish, 4.4 percent French, and under 2 percent for most other languages.[1] This imbalance produces systematic performance gaps. Lai et al. (2023) find ChatGPT achieves 85 percent accuracy on English reasoning tasks versus 65 percent in Chinese—a 20 percentage point gap. Ahuja et al. (2023) document that GPT-3.5 performance in low-resource languages lags English by 30–40 percent across logical reasoning, mathematics, and commonsense tasks. These gaps reflect data scarcity rather than linguistic structure: models simply have less statistical evidence from which to learn in underrepresented languages.

The cognitive cost of translation. Translation does not solve the language barrier. Philosophers from Heidegger to Wittgenstein argue that language constitutes thought itself—the mother tongue is the cognitive architecture within which we reason and solve problems. When we think spontaneously, we think in our native language; this "inner voice" guides cognition at its most fundamental level. Translation disrupts this flow. A Tokyo

[1] OpenAI's technical report for GPT-3 confirms that Common Crawl constitutes the majority of training data (Brown et al., 2020). Bender et al. (2021) and Dodge et al. (2021) document its dominance across major models and its systematic overrepresentation of English-language content.

software engineer debugging code thinks naturally in Japanese but must translate the problem into English to use ChatGPT, evaluate the English response, and translate back before implementing solutions. Each step imposes cognitive costs—not just time but mental effort required to map concepts across linguistic frameworks. For AI applications requiring rapid iteration, accumulated friction from constant translation substantially reduces productivity gains relative to native-language interaction.

Complementarity with knowledge ecosystems. English dominates global technical communication, creating self-reinforcing dynamics that compound linguistic advantages over time. Because knowledge production clusters around English-language platforms—Stack Overflow, GitHub, technical documentation—AI models trained on this corpus inherit English ecosystem structure. As English speakers use AI to generate content, they produce more English material feeding back into future training data. Linguistic advantage today thus compounds advantage tomorrow: the gap between well-represented and underrepresented languages widens not just across users but across model generations.

These three channels aggregate to economy-wide labor productivity divergence. A growing body of evidence documents that AI raises productivity at the task and firm level (Brynjolfsson et al., 2025; Noy and Zhang, 2023; Dell'Acqua et al., 2023). Our framework posits that these gains materialize primarily when linguistic compatibility is high. When linguistic barriers force constant translation and cognitive friction, these gains shrink or disappear entirely. We now formalize this mechanism in a structural model that generates precise, testable predictions.

2.2 Model Environment

We build on the technology adoption framework of Acemoglu et al. (2026), henceforth AAJ, extending it to incorporate AI as a distinct adoption channel whose effectiveness depends on linguistic compatibility with training data. Full derivations are provided in Appendix C.

Consider a world economy consisting of a continuum of countries indexed by i. Each

country produces a final good by combining a set of intermediate inputs with an elasticity of substitution $\sigma > 1$:

$$Y_{it} = Q_{it}^{\frac{\sigma-2}{\sigma-1}} \left(\int_0^{Q_{it}} y_{jt}^{\frac{\sigma-1}{\sigma}} \, dj \right)^{\frac{\sigma}{\sigma-1}}, \tag{1}$$

where Q_{it} denotes the measure of intermediate varieties available in country i at time t and y_{jt} is the quantity of intermediate j. Each intermediate is produced one-for-one from labor by a monopolistic firm. As in AAJ, symmetric equilibrium pricing and labor market clearing yield the aggregate relationship $Y_{it} = Q_{it}$: output per worker equals the number of intermediate varieties. A country grows richer only by mastering new production techniques from the world technology frontier Q_t^W, which grows at a constant exogenous rate $g_W > 0$. We define country i's relative technology position as $q_{it} \equiv Q_{it}/Q_t^W \in (0, 1]$.

A representative household in each country has logarithmic preferences with discount rate $\rho > 0$. Per-intermediate monopoly profit is $\pi = 1/\sigma$, of which a fraction β_i survives institutional expropriation.

2.3 Technology Adoption with AI

A firm in country i can invest $\Gamma \cdot Z$ units of the final good to obtain a flow rate of new intermediates $B_i \cdot Z$. Our key departure from AAJ is to decompose the arrival rate into two distinct channels:

$$B_i = \theta_i \left[\underbrace{q_i^{-\varepsilon}}_{\text{conventional}} + \underbrace{\lambda_i \cdot \mu \cdot q_i^{-\varepsilon}}_{\text{AI-augmented}} \right] = \theta_i \left(1 + \lambda_i \mu\right) q_i^{-\varepsilon}, \tag{2}$$

where:

- $q_i^{-\varepsilon}$ is the *conventional* adoption channel, representing the standard Gerschenkron (1962) mechanism: countries further from the frontier have more technologies available to imitate;
- $\mu \geq 0$ parameterizes AI's productivity enhancement ($\mu = 0$ before ChatGPT);

- $\lambda_i \in [0, 1]$ is country i's *linguistic compatibility* with AI training data—the population-weighted overlap between domestic languages and the language composition of AI training corpora, corresponding directly to our empirical Language Intensity variable;
- $\theta_i > 0$ is absorptive capacity, encompassing human capital, management practices, and all other non-linguistic determinants of a country's ability to assimilate frontier knowledge.

Three features of this specification warrant discussion. First, the multiplicative structure—λ_i enters as a coefficient on the AI channel rather than additively shifting θ_i—captures the idea that linguistic compatibility is *specific to AI*. A Japanese engineer's conventional R&D capacity is unaffected by English-language training data; only her AI-assisted adoption depends on λ. Second, both channels share the same distance-to-frontier elasticity ε, reflecting our assumption that AI accelerates the same underlying process of knowledge absorption rather than introducing a qualitatively different mechanism. Third, the model nests AAJ exactly: when $\mu = 0$, the AI channel vanishes and the standard framework is recovered. ChatGPT's release corresponds to a discrete, unanticipated shift from $\mu = 0$ to $\mu > 0$.

This last point connects to AAJ's own discussion of AI. They explicitly warn that AI, optimized for advanced economies, may prove "inappropriate" for developing countries, exacerbating global inequality—but leave the mechanism unspecified. Our extension provides the missing specification: the dimension of inappropriateness is linguistic, not capital- or skill-based.

2.4 Steady State

The equilibrium structure follows AAJ closely. Define effective absorptive capacity as $\tilde{\theta}_i \equiv \theta_i(1 + \lambda_i\mu)$. The value function for a firm controlling n intermediates satisfies a standard Hamilton-Jacobi-Bellman equation, and the first-order condition for optimal adoption effort

pins down the per-intermediate value:

$$v_i = \frac{\Gamma}{\tilde{\theta}_i \, q_i^{-\varepsilon}}. \tag{3}$$

Combining with the asset pricing condition $v_i = \beta_i \pi/(\rho + g_W)$ yields the balanced growth path (BGP) relative technology:

$$q_i^* = \left(\frac{\theta_i(1 + \lambda_i \mu) \cdot \beta_i \pi}{\Gamma(\rho + g_W)} \right)^{1/\varepsilon}. \tag{4}$$

On the BGP, each country grows at the frontier rate g_W but occupies a distinct position q_i^* determined by its fundamentals. When $\mu = 0$, this reduces to the standard AAJ result:

$$q_i^{*,\text{pre}} = \left(\frac{\theta_i \beta_i \pi}{\Gamma(\rho + g_W)} \right)^{1/\varepsilon}, \tag{5}$$

which depends only on absorptive capacity θ_i and institutional quality β_i. Language plays no role.

When AI arrives ($\mu > 0$), the ratio of post- to pre-AI steady states is:

$$\frac{q_i^{*,\text{post}}}{q_i^{*,\text{pre}}} = (1 + \lambda_i \mu)^{1/\varepsilon}. \tag{6}$$

The full derivation, which follows AAJ closely with θ_i replaced by $\tilde{\theta}_i$, is provided in Appendix C.This expression is the model's central result. The AI-driven productivity gain depends *only* on linguistic compatibility λ_i and AI productivity μ. It is independent of absorptive capacity θ_i, institutional quality β_i, and all other traditional determinants of economic development. We formalize the implications in three propositions.

2.5 Theoretical Predictions

Proposition 1 (Language as an independent determinant) *The proportional gain in steady-state technology from AI, $(1+\lambda_i\mu)^{1/\varepsilon}$, depends only on linguistic compatibility λ_i and AI productivity μ. It is independent of absorptive capacity θ_i, institutional quality β_i, and all other traditional determinants of economic development.*

Proposition 1 provides the theoretical counterpart to our baseline empirical result: Language Intensity predicts AI-driven productivity growth after controlling for GDP per capita, human capital, and internet penetration (Table 2). The shock from $\mu = 0$ to $\mu > 0$ leaves high-λ economies below their new steady state, generating the post-2022 acceleration we test in the event study (Figure 3).

Proposition 2 (Language as a necessary condition) *When $\lambda_i = 0$, $q_i^{*,\text{post}} = q_i^{*,\text{pre}}$ regardless of θ_i and β_i. Linguistic compatibility is a necessary condition for capturing any productivity gain from AI. No level of absorptive capacity or institutional quality can substitute for it.*

This result is qualitatively stronger than standard complementarity. In conventional models of appropriate technology (Acemoglu et al., 2026; Moscona and Sastry, 2025), a mismatch between frontier technology and local conditions reduces the gains from adoption but does not eliminate them entirely. Here, the AI channel is completely shut off when $\lambda_i = 0$: the technology's effectiveness depends on a linguistic interface that cannot be bypassed through investment or institutional reform. Japan illustrates this starkly: despite possessing among the world's highest levels of absorptive capacity and institutional quality, its near-zero linguistic compatibility with English-dominated AI training data renders the AI channel inoperative.

Proposition 3 (Linguistic leapfrogging) *Consider two countries R (rich) and P (poor)*

with $\theta_R\beta_R > \theta_P\beta_P$ (so that $q_R^{*,\text{pre}} > q_P^{*,\text{pre}}$) but $\lambda_R < \lambda_P$. There exists a threshold

$$\bar{\mu} = \frac{\theta_R\beta_R - \theta_P\beta_P}{\theta_P\beta_P\lambda_P - \theta_R\beta_R\lambda_R} \tag{7}$$

such that for all $\mu > \bar{\mu}$, $q_P^{*,\text{post}} > q_R^{*,\text{post}}$: the poor English-speaking country overtakes the wealthy non-English country in steady-state productivity.

Proof. In the pre-AI steady state, $q_R^{*,\text{pre}} > q_P^{*,\text{pre}}$ since $\theta_R\beta_R > \theta_P\beta_P$. In the post-AI steady state, $q_P^{*,\text{post}} > q_R^{*,\text{post}}$ if and only if $\theta_P\beta_P(1+\lambda_P\mu) > \theta_R\beta_R(1+\lambda_R\mu)$. Rearranging yields the condition $\mu > \bar{\mu}$ given in (7). A finite positive solution exists whenever $\theta_P\beta_P\lambda_P > \theta_R\beta_R\lambda_R$—that is, when the poor country's linguistic advantage is sufficiently large relative to the rich country's traditional advantage. This condition holds for pairs such as Nigeria ($\lambda \approx 1.00$) versus Japan ($\lambda \approx 0.11$). ■

The threshold $\bar{\mu}$ is lower when (i) the linguistic gap $\lambda_P - \lambda_R$ is larger, (ii) the initial income gap is smaller, or (iii) the poor country's linguistic advantage is more extreme. As AI becomes more central to production (μ rises over time), leapfrogging extends to a progressively wider set of country pairs.

2.6 AI and Global Inequality

The model provides an analytical decomposition of how AI reshapes global inequality. Taking logs of (4), the cross-country variance of log steady-state productivity is:

$$\underbrace{\text{Var}\big(\ln q_i^{*,\text{post}}\big)}_{\text{post-AI inequality}} = \underbrace{\text{Var}(\ln q_i^{*,\text{pre}})}_{\text{pre-AI inequality}} + \frac{1}{\varepsilon^2}\underbrace{\text{Var}(\ln(1+\lambda_i\mu))}_{\text{linguistic inequality}} + \frac{2}{\varepsilon}\underbrace{\text{Cov}(\ln q_i^{*,\text{pre}},\, \ln(1+\lambda_i\mu))}_{\text{realignment term}}. \tag{8}$$

The second term is unambiguously positive: AI introduces a new source of cross-country inequality driven entirely by variation in linguistic compatibility. The third term depends on the correlation between pre-existing income and linguistic compatibility. Empirically, this correlation is weak ($r \approx -0.09$ in our sample), because English is widely spoken in both

wealthy economies (United States, United Kingdom) and poor ones (Nigeria, Ghana, Kenya), while many wealthy economies have low English prevalence (Japan, South Korea, France). A near-zero covariance term implies that AI *unconditionally* increases global inequality, and that the new inequality it introduces is nearly orthogonal to the pre-existing income distribution.

This is the formal content of our paper's title: generative AI acts as an "inequality amplifier" by adding a new axis of global stratification—linguistic compatibility with training data—that cuts across the traditional rich–poor divide. Unlike previous sources of cross-country inequality, this new dimension cannot be addressed through capital accumulation, institutional reform, or educational investment. It is encoded in the architecture of the technology itself.

3 Data and Measurement

3.1 Productivity Data

We measure productivity using annual labor productivity growth rates from the International Labour Organization's ILOSTAT database, which provides harmonized statistics across countries by dividing real GDP (constant prices, PPP) by total employment. Our sample covers 169 economies from 2015 to 2024, yielding 1,509 country-year observations after excluding 2020. The COVID-19 pandemic generated productivity shocks of heterogeneous sign and magnitude across economies, driven by differences in sectoral composition, lockdown stringency, and fiscal capacity. These cross-country differences are unrelated to linguistic compatibility, and including 2020 would therefore introduce spurious variation that could contaminate identification, particularly by obscuring the parallel pre-trends that our difference-in-differences design requires. Appendix Table A1 reports results including 2020; the baseline coefficient remains stable, though pre-trend tests are less clean, consistent with this concern. The pre-treatment period (2015–2021, excluding 2020) exhibits mean produc-

tivity growth of 1.62 percent; the post-ChatGPT period (2022–2024) shows mean growth of 1.54 percent. We use growth rates rather than levels because difference-in-differences estimation identifies treatment effects through within-country changes around the shock date; level differences are absorbed by country fixed effects.

3.2 Language Intensity Construction

Language Intensity measures linguistic compatibility between a country's population and AI training data. This is our core explanatory variable and merits detailed explanation.

We obtain language usage data from the CIA World Factbook, which reports the percentage of each country's population speaking different languages, covering languages spoken by at least 5 percent of the population. This is a standard data source in the international economics literature on language and trade: Melitz and Toubal (2014), in their widely cited construction of bilateral linguistic variables, rely on the same source for language data across 195 countries. For AI training corpus composition, we use Common Crawl statistics. Common Crawl is a nonprofit that archives the public web and produces openly available datasets widely used to train large language models. OpenAI has confirmed GPT models rely heavily on Common Crawl text. This choice is well-grounded in the computer science literature. Bender et al. (2021) document that large language models suffer from "data documentation debt," with training corpora overrepresenting high-resource languages from wealthy, digitally connected regions. Joshi et al. (2020) show over 60 languages lack sufficient digital text for effective model training. Using Common Crawl composition to proxy AI training data thus reflects actual linguistic biases embedded in generative AI systems.

Language Intensity for country c is constructed as:

$$\text{Language Intensity}_c = \sum_l (\text{Domestic Share}_{c,l} \times \text{Crawl Share}_l) \tag{9}$$

where $\text{Domestic Share}_{c,l}$ is the fraction of country c's population speaking language l, and

Crawl Share$_l$ is the fraction of Common Crawl corpus in language l. To illustrate, consider Singapore versus Japan. Singapore has English (48.3%), Mandarin, Malay, and Tamil: $\text{Intensity}_{\text{Singapore}} = (0.483 \times 0.443) + \ldots \approx 0.237$. Japan has Japanese (100%) and negligible English: $\text{Intensity}_{\text{Japan}} = (1.00 \times 0.047) \approx 0.047$. Singapore's Language Intensity is five times Japan's, despite comparable levels of economic development. This captures our key insight: linguistic compatibility with AI training data cuts across traditional income hierarchies.

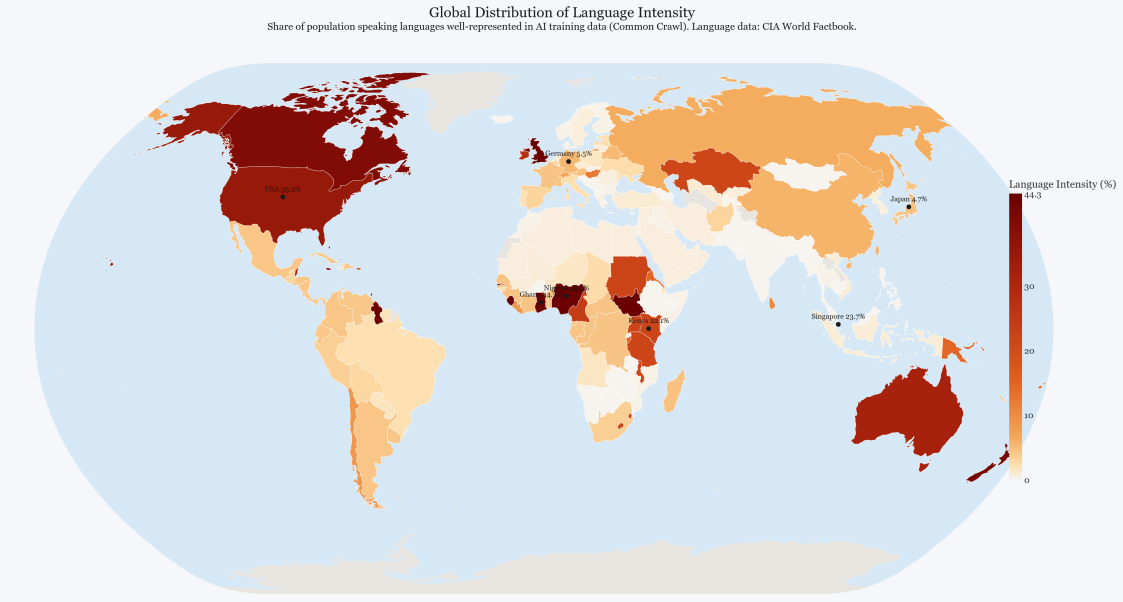


Figure 1: The Map of Language Intensity

Notes: Language Intensity is constructed as the population-weighted average of Common Crawl corpus shares across languages spoken in each economy. Data sources: CIA World Factbook (language usage) and Common Crawl statistics (corpus composition).

The measure captures the probability that a randomly selected individual speaks a language well-represented in AI corpora. It is not simply an English indicator—multilingual countries (Singapore, South Africa) score highly due to substantial English-speaking populations, while monolingual Anglophone countries (US, UK) also score at the maximum. We standardize Language Intensity to mean zero and standard deviation one, allowing coefficients to be interpreted as effects of a one-standard-deviation increase in linguistic compatibility. For robustness, we construct three alternative measures. Method 1 (baseline) uses all languages spoken by at least 5 percent of the population, weighted by Common

Crawl shares—our preferred measure. Method 2 uses only each country's primary language: Primary Intensity$_c$ = Crawl Share(Primary Language$_c$). Method 3 uses only domestic English share: English Intensity$_c$ = Domestic English Share$_c$. Table 2 reports results using all three measures; coefficients remain positive and significant across specifications.

Importantly, the correlation between Language Intensity and log GDP per capita is nearly zero (-0.09), indicating that linguistic compatibility is essentially orthogonal to economic development. Wealthy non-English economies (Japan, South Korea, Germany) have low Language Intensity despite high incomes, while poorer English-speaking economies (Nigeria, Ghana, Kenya) have high Language Intensity despite low incomes. This orthogonality is central to our identification strategy and enables the "linguistic leapfrogging" patterns documented in Section 7.

3.3 Summary Statistics

Table 1: Summary Statistics

	N	Mean	SD	Min	Max
Productivity Growth (%)	1,509	1.536	4.569	-22.02	68.59
Language Intensity (%)	1,509	7.623	12.907	0.00	44.29
Primary Language Intensity (%)	1,509	6.412	12.737	0.00	44.29
English Share Intensity (%)	1,509	5.176	13.109	0.00	44.29
ln(GDP per capita, PPP)	1,509	11.728	2.057	6.49	17.33
Internet Penetration (0-1)	1,509	0.613	0.282	0.01	1.00

Notes: Summary statistics for main variables. Sample: 169 economies, 2015-2024 (year 2020 excluded), 1,509 observations. Productivity Growth: annual labor productivity growth rate (pp). Language Intensity measures: comprehensive (all languages $\geq$5%), primary only, English only. All intensity measures weighted by representation in AI training data (Common Crawl corpus).

Table 1 reports descriptive statistics. Our sample comprises 1,509 country-year observations covering 169 economies from 2015 to 2024 (excluding 2020). Productivity growth averages 1.54 percent (standard deviation 4.57, ranging from -22 to $+69$ percent); the wide range reflects both genuine volatility and measurement challenges in some developing economies, which our clustered standard errors and fixed effects accommodate. Lan-

guage Intensity ranges from effectively zero to 44.3 percent (mean 7.6, standard deviation 12.9) and enters the regressions standardized to mean zero and unit variance. Crucially for identification, income and Language Intensity overlap substantially—many low-intensity economies are wealthy (Japan, South Korea) and many high-intensity economies poor (Nigeria, Ghana)—so the language effect is identified across the full income spectrum rather than confounded with it.

We include log GDP per capita (PPP, constant 2017 international dollars) and internet penetration as time-varying controls for economic development and digital infrastructure; country fixed effects absorb time-invariant characteristics such as institutions, culture, and geography, while year fixed effects absorb global shocks. Section 4 details the specification.

4 Empirical Strategy

4.1 Identification: ChatGPT as a Natural Experiment

ChatGPT's November 30, 2022 release provides an ideal natural experiment. While OpenAI had released GPT-3 via API in 2020, ChatGPT represented a qualitative leap in public engagement. GPT-3 required technical knowledge; ChatGPT offered a free, simple web interface. Within five days, ChatGPT reached one million users. Within two months, it surpassed 100 million users—the fastest technology adoption in history. This explosive uptake was unanticipated even within the AI community. Crucially, the release timing was determined by OpenAI's corporate roadmap, not by economic conditions in any country—satisfying the exogeneity requirement for natural experiments.

Figure 2 presents Google Trends data on global search interest for "AI" and "ChatGPT" from 2019 to 2025. The pattern validates our identification strategy. Throughout 2019–2022, search interest in "AI" remains low and flat, fluctuating around a baseline index of 10–15. At ChatGPT's November 30, 2022 release, search interest spikes vertically. The "ChatGPT" series jumps immediately to peak levels, while the "AI" series rises in parallel, reflecting

broader awareness of generative AI. Both remain elevated through 2025, indicating sustained rather than transient interest. Critically, there is no anticipatory rise before November 2022—the increase is immediate and discontinuous.

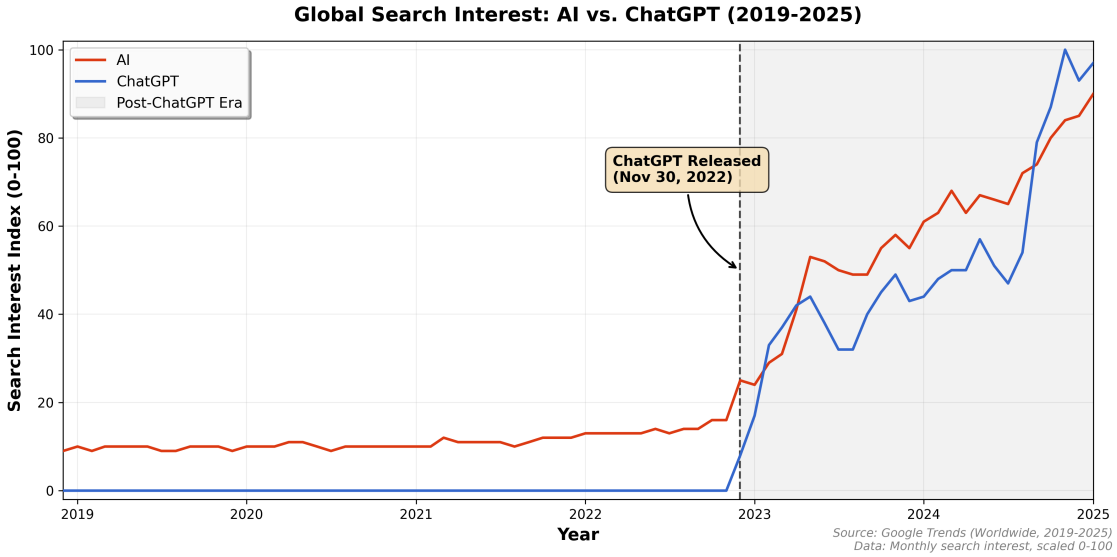


Figure 2: Google Trends Chart of AI and ChatGPT

This evidence validates two features essential for difference-in-differences estimation. First, ChatGPT's release constituted a discrete, identifiable shock rather than gradual diffusion—the vertical spike contrasts sharply with typical technology adoption curves. Second, the shock was unanticipated: the flat pre-period indicates economies were not differentially preparing for generative AI before the release. These features—a sharp, unanticipated discontinuity—are precisely what our empirical design requires to credibly identify causal effects.

4.2 Difference-in-Differences Specification

Our baseline equation is:

$$\Delta\text{Productivity}_{ct} = \beta\left(\text{Language Intensity}_{c} \times \text{Post}_{t}\right) + X'_{ct}\gamma + \alpha_c + \delta_t + \epsilon_{ct} \tag{10}$$

where $\Delta\text{Productivity}_{ct}$ is annual labor productivity growth (%) for country c in year t. The key variable is Language Intensity$_c \times \text{Post}_t$, where $\text{Post}_t = \mathbb{1}[\text{year} \geq 2022]$. The coefficient β captures differential productivity growth for high versus low Language Intensity economies after ChatGPT's release. We include controls X_{ct} (log GDP per capita, internet penetration), country fixed effects α_c (absorbing time-invariant differences), year fixed effects δ_t (absorbing global shocks), and cluster standard errors at the country level.

The coefficient β is directly interpretable. Our baseline estimate $\beta = 0.938$ (Table 2, Column 2) implies a one-standard-deviation increase in Language Intensity raises post-ChatGPT productivity growth by 0.94 percent. Relative to the pre-treatment mean of 1.62 percent, this represents a 58 percent acceleration. Over three years (2022–2024), this compounds to nearly a 3 percentage point cumulative gap between economies one standard deviation apart in Language Intensity.

4.3 Identification Assumptions

Difference-in-differences estimation requires three assumptions. **Parallel trends**: absent ChatGPT, productivity growth trends would have evolved similarly across Language Intensity levels. Event studies in Section 5.2 show flat pre-trends from 2015–2021 and a sharp break in 2022, validating this assumption. **Treatment timing exogeneity**: ChatGPT's release was unrelated to differential economic conditions across language groups. This is plausible given OpenAI's corporate decision-making process. **No confounding**: Language Intensity does not merely proxy for institutions, culture, or other omitted factors. We address this through controls, colonial history analysis (Section 6.1), and AI adoption controls (Section 6.2).

4.4 Potential Threats to Identification

Three threats merit attention; each is addressed in Section 6. First, language might proxy for institutions inherited from British colonization. Our colonial history analysis compares

British, French, and Spanish ex-colonies: only English-speaking former colonies show gains, mirroring training data composition rather than institutional legacies. Second, high Language Intensity economies might simply adopt AI more intensively. We control for Google Trends search intensity; the Language Intensity coefficient remains unchanged, indicating differential effectiveness rather than adoption. Third, pre-existing differential trends could violate parallel trends. Event studies show perfectly flat pre-trends, ruling out this concern. Section 6 presents detailed tests; together, these establish that linguistic compatibility with AI training data causally affects productivity divergence.

5 Main Results

5.1 Baseline Effect: Language Matters

We begin by estimating how linguistic compatibility with AI training data shapes productivity growth after ChatGPT's release.

Table 2 presents our baseline difference-in-differences estimates across six specifications. We estimate three specifications of Language Intensity to ensure our findings do not depend on a particular measurement choice. The first two columns use our preferred measure—a comprehensive weighted index combining all languages spoken by at least 5 percent of the population with their representation in Common Crawl training data. Columns 3-4 employ a simplified approach using only each economy's primary language. Columns 5-6 use an even simpler measure: each economy's English-speaking population share. For each measure, we report results without controls (odd columns) and with controls for log GDP per capita and internet penetration (even columns). Across all specifications, the coefficient on Language Intensity $\times$ Post is positive and statistically significant, indicating our core finding is robust to measurement choices.

Our preferred specification (Column 2) combines the comprehensive Language Intensity measure with economic controls. The coefficient on Language Intensity $\times$ Post is stable

across specifications—1.120 without controls (Column 1) and 0.938 with controls (Column 2)—suggesting our results are not driven by omitted economic or technological factors.

Table 2: Baseline DID Results

	Productivity Growth					
	Lang. Intensity		Primary Lang.		English Only	
	(1)	(2)	(3)	(4)	(5)	(6)
Lang. Int. × Post	1.120*	0.938***				
	(0.587)	(0.360)				
Primary × Post			1.150*	0.948***		
			(0.606)	(0.361)		
English × Post					1.072*	0.850**
					(0.610)	(0.355)
ln(GDP pc)		10.123*		10.025*		10.036*
		(5.353)		(5.325)		(5.375)
Internet		-3.218		-3.322		-3.397
		(2.324)		(2.317)		(2.328)
Observations	1,509	1,509	1,509	1,509	1,509	1,509
Adjusted R^2	0.220	0.252	0.221	0.253	0.219	0.250
Pre Period mean	1.616	1.616	1.616	1.616	1.616	1.616
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variable: annual labor productivity growth (pp). All language intensity measures standardized to mean 0, SD 1. Lang. Intensity: all languages ≥5% weighted by training data. Primary Lang.: primary language only. English Only: English share only. Post=1 for 2022+. Year 2020 excluded. Controls: ln(GDP per capita), Internet penetration. SE clustered at country level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

The effect is economically substantial. A one-standard-deviation increase in Language Intensity raises post-ChatGPT productivity growth by 0.94 percent. To contextualize this magnitude, the mean productivity growth rate in the pre-ChatGPT period (2015–2021, excluding 2020) was 1.62 percent. Thus, a one-standard-deviation increase in linguistic compatibility raises growth by 58 percent relative to the baseline rate—a large effect by any standard. In concrete terms, consider two high-income Asian economies with similar digital infrastructure but divergent linguistic compatibility: Singapore (Language Intensity = 23.7%) and Japan (Language Intensity = 4.7%). Our estimates indicate that Singapore's productivity growth exceeds nearly 1 percent higher per year than Japan's in the post-

ChatGPT period, holding other factors constant. The AI era is still in its early stages, yet should this gap in growth rates persist, the two economies may gradually diverge in ways that income or technological capacity alone cannot explain. Language compatibility with AI training data represents a genuinely new axis of inequality, one whose full consequences are only beginning to unfold. The more striking pattern, whereby low-income English-speaking economies outperform wealthy non-English ones, is taken up in Section 7.

Results are robust to alternative constructions of Language Intensity: coefficients remain positive and significant whether using only each economy's primary language (Column 4) or English-speaking population share alone (Column 6). The comprehensive measure yields the largest coefficient, likely capturing the full linguistic diversity of multilingual economies that simpler measures miss. These baseline results are robust to controls for differential AI adoption intensity across economies, as we demonstrate in Section 6.2. We first validate the causal interpretation through event study specifications.

5.2 Dynamic Effects: Validating Parallel Trends

The difference-in-differences estimator requires that, absent ChatGPT's release, productivity growth would have evolved similarly across economies with different Language Intensity levels. We test this using event study specifications that estimate year-by-year effects relative to a base year of 2021, omitting 2020 to avoid COVID-19 disruptions.[2] Figure 3 plots the estimated coefficients and 95 percent confidence intervals.

The pre-treatment pattern strongly supports parallel trends. From 2015 through 2019, all coefficients hover near zero with confidence intervals spanning zero, with no discernible trend. This confirms that high- and low-Language-Intensity economies followed similar productivity trajectories before ChatGPT's release.

Following ChatGPT's November 2022 release, a sharp divergence emerges and persists

[2]Including 2020 produces less clean patterns due to COVID-19's heterogeneous effects across economies; however, pre-treatment coefficients remain statistically indistinguishable from zero and the post-treatment divergence persists. Results are reported in Appendix Figure A1.

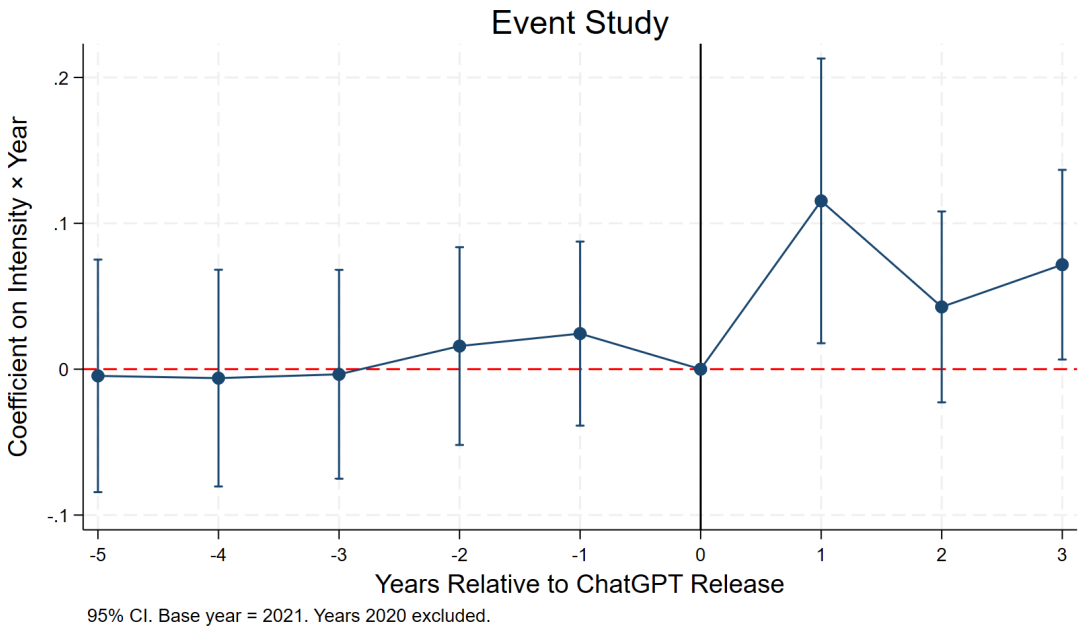


Figure 3: Event Study: Year-by-Year Effects of Language Intensity on Productivity Growth
Notes: Coefficients from event study specification relative to base year 2021. Year 2020 excluded. Dependent variable is annual labor productivity growth (%). 95 % confidence intervals based on standard errors clustered at the country level.

through 2024. The post-period coefficients remain consistently positive and elevated relative to the pre-period baseline, with some year-to-year variation likely reflecting business cycle fluctuations and the gradual diffusion of AI adoption across firms and sectors. Critically, the divergence materializes precisely in 2022 and does not anticipate the treatment, reinforcing a causal interpretation.

Together with the baseline results in Section 5.1, these findings establish two core points: Language Intensity significantly predicts AI-driven productivity gains, and this relationship is causal, validated by parallel pre-trends and precise timing around ChatGPT's release.

6 Robustness Checks

6.1 Addressing Institutional Confounds: Colonial History

A primary concern is that Language Intensity may proxy for institutional quality, cultural factors, or other historical legacies correlated with economic performance. English-speaking economies, for instance, are often former British colonies, which Acemoglu et al. (2001) show inherited relatively strong property rights and legal institutions. If our results merely reflect these institutional advantages rather than linguistic compatibility with AI training data, the interpretation would be fundamentally different. We address this concern using colonial history to generate quasi-exogenous variation in language exposure while holding institutional legacies relatively constant.

Our strategy exploits two sources of identification. First, we restrict the sample to former colonies, comparing economies with shared colonial experiences but different colonizer languages. This isolates linguistic variation from broader institutional differences between colonized and non-colonized nations. Second, we leverage the identity of the colonizing power as a source of plausibly exogenous variation in language exposure: whether a country speaks English, French, or Spanish today largely reflects which European power colonized it centuries ago—a historical accident unrelated to modern AI technology. That former British colonies exhibit larger effects reflects the historical accident of English dominance in AI training corpora, not any beneficial legacy of colonialism; we return to this distinction in Section 10. Table 3 presents results under both approaches.

Column 1 restricts the sample to former British colonies, exploiting within-group variation in Language Intensity. Even among economies sharing British institutional heritage—Kenya, Ghana, Nigeria—a one-standard-deviation increase in linguistic compatibility raises productivity growth by 1.35 percent, confirming that language matters independently of colonial legacy.

Column 2 expands the sample to all former colonies, yielding a similar effect of 1.24

Table 3: Colonial History and Language Effects

	Productivity Growth		
	(1) British	(2) All Colonies	(3) By Type
Language Intensity × Post	1.351***	1.238***	
	(0.460)	(0.435)	
Lang × Post × British Colony			1.340***
			(0.469)
Lang × Post × French Colony			-0.027
			(0.900)
Lang × Post × Spanish Colony			-0.196
			(2.238)
Log GDP per capita	20.329***	17.083***	17.024***
	(3.088)	(4.043)	(4.094)
Internet penetration (%)	0.855	2.037	1.658
	(5.979)	(4.004)	(4.072)
Observations	268	673	673
Adjusted R^2	0.632	0.400	0.396
Country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Notes: Dependent variable is annual labor productivity growth rate (percentage points). Column 1 restricts sample to former British colonies to control for institutional similarities. Column 2 includes all former colonies (British, French, Spanish). Column 3 separately estimates effects by colonial type.

All specifications include country and year fixed effects. Standard errors clustered at country level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

percent. The consistency across these two specifications suggests our baseline finding is not an artifact of institutional differences between colonized and non-colonized economies.

Column 3 provides the most direct evidence. Effects concentrate entirely among former British colonies, while French and Spanish colonies show no significant gains. This pattern mirrors AI training data composition—not colonial institutional legacies: Common Crawl comprises 44.3 percent English but only 4.4 percent French and 4.3 percent Spanish. Since institutional structures are broadly similar across colonial types, the differential effect can only be attributed to language.

These results strengthen the causal interpretation of our baseline findings. By restricting analysis to former colonies—where institutional and cultural legacies are broadly shared—we isolate linguistic compatibility as the operative mechanism. The sharper test in Column 3 reinforces this conclusion: the productivity effect tracks the language composition of AI training data, not the identity of the colonizer per se. Institutional quality, legal heritage, and cultural legacies cannot account for a pattern that so precisely mirrors the English dominance of Common Crawl.

6.2 Adoption versus Effectiveness: Controlling for AI Search Intensity

A natural concern is that economies with high Language Intensity may simply adopt AI more intensively, rather than use it more effectively. If our results reflect differential awareness or uptake—populations in English-speaking countries searching for, downloading, or experimenting with AI tools at higher rates—then the interpretation shifts from one about linguistic barriers to effectiveness toward one about diffusion rates. These mechanisms carry distinct policy implications: the former calls for multilingual AI development, while the latter points toward digital literacy and technology access programs.

We address this concern in two steps using Google Trends data on search intensity for the term "AI" across economies and over time. Google search data provide a revealed-preference

measure of AI awareness and interest that is available at global scale and comparable frequency, unlike firm-level adoption surveys. We convert relative search indices into absolute search volumes using Glimpse API anchor data, ensuring cross-country comparability, and standardize the resulting measure to mean zero and unit variance. Table 4 presents results.

Table 4: Language Intensity and AI Adoption: Orthogonality and Robustness

	Productivity Growth		**AI Search Volume (Std)**	
	(1) Baseline	(2) AI Control	(3) Panel	(4) Cross-Section
Language Intensity × Post	0.938***	0.955***		
	(0.360)	(0.360)		
AI Search Intensity (Std)		−0.132		
		(0.121)		
Language Intensity			0.022	0.021
			(0.030)	(0.058)
Log GDP per capita	10.123*	10.087*		
	(5.353)	(5.392)		
Internet penetration (%)	−3.218	−3.004		
	(2.324)	(2.357)		
Observations	1,509	1,455	1,455	163
R^2 / Adj. R^2	0.252	0.255	0.000	0.001
Country FE	Yes	Yes		
Year FE	Yes	Yes	Yes	

Notes: Columns 1–2 dependent variable: annual labor productivity growth rate (percentage points). Columns 3–4 dependent variable: standardized AI search volume (Google Trends). The Language Intensity coefficient is virtually unchanged after controlling for AI search intensity (0.955 vs. 0.938), reflecting differential effectiveness, not differential adoption. Columns 3–4 confirm that Language Intensity does not predict AI search volume, establishing orthogonality between linguistic compatibility and AI adoption.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

We first examine directly whether Language Intensity predicts AI search volume. Columns 3–4 regress standardized AI search intensity on Language Intensity, using both the country-year panel (Column 3, with year fixed effects) and a cross-sectional collapse to country means (Column 4). In both specifications, the coefficient on Language Intensity is near zero (0.022) and statistically insignificant, with R^2 values below 0.01. Linguistic compatibility with AI

training data bears no systematic relationship to how intensively populations search for AI-related information. The two dimensions—language match and adoption intensity—are orthogonal.

We then ask whether controlling for AI search intensity alters the baseline productivity result. Column 2 adds standardized AI search volume as a control alongside our baseline specification in Column 1. The Language Intensity coefficient moves from 0.938 to 0.955, remaining statistically significant at the 1 percent level with nearly identical standard errors. The AI search intensity control itself enters negatively (-0.132) and insignificantly, indicating that awareness of and curiosity about AI do not independently predict productivity gains.

Taken together, these results rule out differential adoption as an explanation for our findings. Language Intensity does not predict AI search volume (Columns 3–4), and controlling for search volume leaves the Language Intensity effect unchanged (Columns 1–2). Populations may search for AI at similar rates, but those whose languages align with AI training corpora extract substantially larger productivity benefits. Combined with the colonial history analysis in Section 6.1, these tests address the two most plausible alternative interpretations of our baseline results: institutional quality is ruled out by colonial comparisons, and differential adoption is ruled out here. What remains is the mechanism we propose: linguistic compatibility with AI training data determines how effectively economies leverage generative AI for productivity growth.

7 Linguistic Leapfrogging: Winners and Losers

We have shown that linguistic compatibility causally predicts AI-driven productivity growth and that the relationship is robust to institutional and adoption-based explanations. The next question is more granular: who are the winners and losers?

Traditional models of technology diffusion offer competing predictions. Comin and Hobijn (2010) and Comin and Mestieri (2018) document that wealthy nations systematically adopt

new technologies earlier, driving income divergence. The Gerschenkron (1962) tradition suggests the opposite: laggard economies enjoy an advantage of backwardness, leapfrogging outdated technologies by adopting frontier ones directly. Acemoglu et al. (2026) extend this concern, arguing that technologies optimized for advanced economies may fail to transfer to developing ones. Our findings reveal a different pattern altogether: what determines who benefits from AI is neither income nor distance from the technological frontier, but linguistic compatibility with training data—a dimension absent from all three traditions.

We document this pattern through two complementary analyses: a winners-and-losers matrix that categorizes economies by income and linguistic compatibility (Section 7.1), and a closer examination of what amplifies the effect and whether it survives demanding institutional controls (Section 7.2).

7.1 The Winners and Losers Matrix

We categorize economies along two dimensions simultaneously: income (above or below median 2019 GDP per capita) and Language Intensity (above or below median). This yields a two-by-two matrix with four groups, using poor low-compatibility economies as the baseline. Table 5 presents the results.

The pattern overturns traditional expectations. Quadrant 1 (rich, high compatibility)—the United States, United Kingdom, Australia—shows productivity gains of 1.86 percent, as expected. These are the predicted winners: wealthy and linguistically aligned. Quadrant 2 (poor, high compatibility)—Ghana, Nigeria, Kenya—experiences gains of 1.26 percent. These are the *unexpected winners*: despite low incomes, high linguistic compatibility generates substantial productivity acceleration. Strikingly, Quadrant 3 (rich, low compatibility)—Japan, South Korea, Germany—shows an effect of -0.01 percent, statistically indistinguishable from zero. These are the *unexpected losers*: wealth, infrastructure, and institutions do not translate into AI-driven gains when linguistic compatibility is absent. Quadrant 4 (poor, low compatibility) forms the baseline, as expected.

Table 5: Winners and Losers: Language Outweighs Wealth in the AI Era

		Language Compatibility with AI	
		HIGH (English-speaking)	**LOW** (Non-English)
GDP per capita	**HIGH** (Rich)	EXPECTED WINNERS **1.861*** (1.039)	UNEXPECTED LOSERS −0.014 (0.633)
	LOW (Poor)	UNEXPECTED WINNERS **1.263*** (0.687)	EXPECTED LOSERS 0.000 *[baseline]*

Notes: Each cell displays the productivity growth effect (percentage points) for countries in that income-language quadrant, relative to the baseline group (poor, non-English countries). Standard errors in parentheses, clustered at country level.
*$p < 0.10$, **$p < 0.05$, ***$p < 0.01$.

The contrast between Quadrants 2 and 3 is stark: poor high-compatibility economies show gains of 1.26 percent, while wealthy low-compatibility economies show essentially zero effect (−0.01%). To grasp the magnitude, consider Nigeria (GDP per capita approximately $2,447) versus Japan ($33,800)—nearly fourteenfold higher. Japan possesses world-class digital infrastructure, universal broadband, highly educated workers, and leading technology firms. Yet Nigeria, with widespread English use, experiences measurably higher AI-driven productivity growth. Poor high-compatibility economies achieve gains statistically indistinguishable from those of wealthy Anglophone nations (Quadrants 2 versus 1). Wealth amplifies linguistic advantages slightly—rich high-compatibility economies gain 1.86 percent versus 1.26 for poor ones—but cannot create them. Without linguistic compatibility, economic resources deliver no discernible AI productivity boost. The modest significance levels ($p<0.10$) likely reflect reduced statistical power from splitting the sample into four quadrants rather than an absence of underlying effects. Of course, the linguistic advantage these

economies enjoy is not a beneficial colonial legacy but a historical contingency in training data composition—one whose inequitable implications we address in Section 10.

Language compatibility is a *necessary condition* for capturing AI gains. Wealth, institutions, and human capital cannot substitute for it. Without it, even the most advanced economies show no productivity response (Quadrant 3). With it, even poor economies experience substantial acceleration (Quadrant 2).

This is linguistic leapfrogging: a novel form of catch-up driven by compatibility with a specific technology's training corpus, operating independently of traditional development pathways. This phenomenon introduces a new axis of stratification into the technology diffusion literature. Prior frameworks organized economies along a single dimension—income or distance from the technological frontier—and predicted that catch-up would follow economic lines. Linguistic leapfrogging reveals a second, orthogonal dimension: linguistic proximity to training data. These two dimensions are largely independent— wealthy economies and poor ones alike can fall on either side of the linguistic divide—which is precisely why the pattern is so striking. Japan and Nigeria are not competing on the same terms; they are being sorted by a criterion that economists had not previously considered.

7.2 What the Pattern Survives: Amplifiers and a Colonial Test

The matrix establishes that linguistic compatibility is necessary. Two further analyses ask what determines how much economies extract from it, and whether the pattern might be institutional rather than linguistic. First, income and digital infrastructure act as amplifiers, not substitutes. Splitting the sample by internet penetration and by income (Table 6), the language effect is positive but negligible in low-internet economies and more than twice as large, and precisely estimated, in high-internet ones; the same holds across the income split, where the effect is roughly three times larger among high-income economies. Neither resource produces gains on its own—each amplifies the gains that linguistic compatibility makes possible. This is the "double divide" of the AI era: language opens the door, while income

and digital infrastructure determine how far one can walk through it. It also reconciles the matrix, where wealthy high-compatibility economies gain somewhat more than poor ones—not because wealth creates the advantage, but because it magnifies what language has already unlocked.

Table 6: Heterogeneity Analysis: Digital Infrastructure and Economic Development

	Internet Penetration		**Income Level**	
	(1) Low	(2) High	(3) Low	(4) High
Language Intensity × Post	0.399	0.839**	0.367	1.095***
	(0.364)	(0.414)	(0.286)	(0.417)
Observations	747	754	739	761
Adjusted R^2	0.304	0.295	0.221	0.338
Country FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: Dependent variable is annual labor productivity growth rate (percentage points). Sample split at median for each dimension: median internet penetration (Columns 1–2) and median 2019 GDP per capita (Columns 3–4). * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Second, the pattern is linguistic, not a colonial-institutional artifact. One might worry that the advantage of poor English-speaking economies reflects institutions inherited from British colonization rather than language itself. Restricting attention to former colonies—which share broadly similar histories of colonial extraction and post-independence state-building—holds those legacies approximately constant and isolates language by comparing British ex-colonies to French and Spanish ones. The leapfrogging pattern not only survives but strengthens: poor British colonies gain 2.55 percent relative to poor non-British ones, more than double the full-sample effect, with the largest gains in the poorest colonies and the four-quadrant ranking unchanged (Appendix F). That the effect grows under tighter institutional controls is precisely what the language hypothesis predicts and what an institutional alternative would struggle to explain. A still cleaner separation—holding every national characteristic fixed and varying only the language of the task—comes from the controlled

experiment of Section 8.

8 Micro-Foundations: Direct Experimental Evidence

Our cross-country analysis establishes that linguistic compatibility shapes AI-driven productivity growth. A natural question is *where* this difference comes from—through what concrete, task-level channel a linguistically compatible economy captures gains that an incompatible one does not. To illuminate this mechanism directly, we complement the macro analysis with a controlled experiment: when a generative-AI tool is put to work on a common business decision, does the economic value it produces—and the cost of obtaining it—depend on the language of the task? The evidence is deliberately narrow in scope: it speaks to the quality and cost of an AI tool's decisions, not to the productivity of any worker, which our data cannot address. But within that scope it is sharp, and it reaches languages—Swahili, Yoruba—for which no firm-level dataset exists, precisely the economies our framework identifies as most exposed.[3]

8.1 Design

Tasks. We build a battery of business decision tasks around *classic, well-studied problem types*—linear programming (product mix), decision analysis under uncertainty (value of information), and market-entry strategy—rather than natural-language-processing benchmarks. The scenarios are original; only the underlying problem structures are canonical. The tasks span a spectrum of decision *structure*: *computational* decisions with a well-defined optimum (production planning and its capacity/shadow-price sensitivity); a *semi-structured* decision under uncertainty (whether to pay for a market study before launching, requiring Bayesian updating and decision-tree fold-back); and *judgment* decisions with no closed-form

[3]The experiment grew out of our work on Zhang et al. (2025), a multilingual agent-based simulation framework whose finding—that native-language simulation tracks real-world behavior more faithfully than English-only simulation—motivated the question we pursue here in an economic setting.

answer (choosing a market-entry strategy consistent with a stated corporate priority; reading the true intent behind indirect business communication). Each is posed as a realistic memorandum in which the relevant quantities and constraints are embedded in prose, so the model must comprehend the situation before acting. Crucially, every task admits a *computable economic benchmark*, so we score outputs by *economic value retained*—the share of the optimal payoff a decision captures—not by linguistic accuracy. Table 7 summarizes the tasks; Appendix D gives full texts.

Table 7: Business decision tasks

Task	Structure	Decision	Computable benchmark
Production planning	Computational	Optimal product mix under capacity and demand limits	Profit vs. LP optimum
Capacity (shadow price)	Computational	Whether to rent extra bottleneck capacity at a stated price	Net profit vs. marginal value
Value of information	Semi-structured	Whether to buy a market study before launching, and what to build	EV-maximizing policy (decision tree + Bayes)
Market-entry strategy	Judgment	Which market fits the firm's stated strategic priority	Uniquely consistent choice
Business communication	Judgment	The intent behind an indirect message, and the right response	Subtext-consistent decision

A representative task. The *value-of-information* problem—a staple of decision analysis—fixes ideas. The model receives a memorandum of the following form (values vary across instances):

> *Capital project memo.* We are weighing whether to launch a new product line: build a LARGE plant, a SMALL plant, or not launch ($0). If demand is HIGH, the large plant clears about $700 and the small one $320; if LOW, the large plant returns −$400 and the small one $180. Marketing puts the chance of high demand at 40%. Before committing, we may pay $70 for a market study; historically it has called genuinely

> strong launches "favourable" 80% of the time, but has also done so for ultimately weak launches 10% of the time. *Which course of action maximizes expected value?* (A) build large now; (B) build small now; (C) study, then build large if "favourable" and small otherwise; (D) study, then build large if "favourable" and abandon otherwise.

Answering requires recovering the structure from prose, updating the demand probability by Bayes' rule on the imperfect signal, folding back the tree, and netting out the study's cost. The optimum is computed exactly, so each answer maps to a precise economic value: a capable model solves it in English (value retained 0.97); the identical decision in Yoruba falls to 0.63.

Languages, models, protocol. We present each task in languages spanning four orders of magnitude in training-data representation—English (44.3% of Common Crawl), Chinese (5.8%), Swahili (0.008%), Yoruba (0.0009%)—using the same corpus shares that construct our macro Language Intensity measure, and evaluate four models: three mainstream Western models (GPT-4o-mini, GPT-4.1-mini, Llama-3.1-70B) and a leading Chinese model (DeepSeek-V3). We follow the experimental standards of the multilingual model-evaluation literature (Zhang et al., 2025): each (model, language, instance) is a stateless, temperature-zero query with no shared history, so a model cannot carry information across languages and the within-model cross-language comparison is clean; the same numerical instances are used in every language, with numerals injected programmatically into translated templates so only the prose differs; optima are computed exactly and a strong model attains them in English, confirming each task is well posed; and we report results across all four models. Large language models retain modest run-to-run variability even at temperature zero, so our conclusions rest on the large, replicated patterns rather than precise per-cell values. Outputs that fail to return a parseable decision are counted as zero value, since an unusable output delivers none. All code, prompts, and seeds are released.

8.2 AI's Economic Value Falls with Resourcedness, Most Steeply for Judgment

Figure 4(a) and Table 8 present the core result, averaged across the three Western models. Value retained declines with resourcedness, but very differently by decision type. On *computational* tasks the decline is modest—from 0.72–0.74 in English and Chinese to 0.40 in Yoruba; on *judgment* tasks it is a collapse—from 0.87–0.92 to 0.19. The contrast is starkest for the most capable model: GPT-4.1-mini is essentially flat on computational tasks across all four languages (1.00 in English and Chinese, 0.95 in Yoruba) yet falls from 1.00 to 0.30 on judgment. Once a capable model extracts the problem, the optimization itself is language-robust; contextual business judgment is not. The pattern is not English-specific: *Chinese*, a non-English high-resource language, retains as much value as English—resourcedness, not English per se, separates winners from losers. And it holds across models: the resourcedness slope on judgment tasks is 0.145, 0.150, and 0.158 for the three Western models, an order of magnitude steeper than for computational tasks.

Table 8: Economic value retained by decision type and language

Decision type	English	Chinese	Swahili	Yoruba
Computational (LP, shadow price)	0.72	0.74	0.62	0.40
Information value (semi-structured)	0.61	0.61	0.51	0.44
Judgment (strategy, communication)	0.87	0.92	0.56	**0.19**

Notes: Value retained $\in [0, 1]$, 1 = optimum; mean across the three Western models, $n = 30$ instances per cell (unparseable outputs counted as zero). Resourcedness slope on judgment tasks ≈ 0.15, consistent across all three models. For the most capable model (GPT-4.1-mini) the computational row is essentially flat ($1.00 \to 0.95$); the average is pulled down by weaker models, sharpening the contrast with judgment.

This collapse is the task-level analogue of Proposition 2, which holds that linguistic compatibility is a *necessary* condition: as $\lambda \to 0$ the AI channel shuts off rather than merely weakening. In Yoruba, value retained on judgment tasks falls to 0.19–0.30, at or below the 0.25 a coin-flip among four options would achieve—at the lowest end of compatibility, the AI tool delivers essentially no usable business judgment. This mirrors the "unexpected losers"

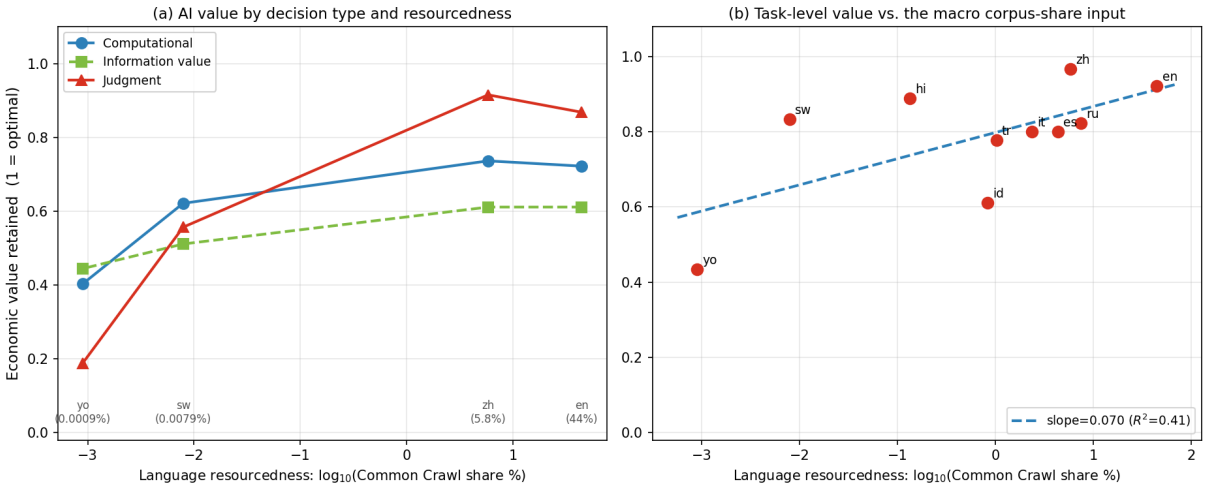


Figure 4: Language resourcedness gates the economic value of AI business decisions
Notes: (a) Economic value retained (1 = optimum) by decision type and language resourcedness ($\log_{10}$ Common Crawl share), averaged across the three Western models. (b) Value retained on the market-entry task for ten languages chosen to span the resourcedness range, against the same corpus share that constructs macro Language Intensity; dashed line OLS ($b = 0.07$, $t = 2.4$, $R^2 = 0.41$). Panel (b) is illustrative; a full multilingual analysis is future work. Unparseable outputs counted as zero value.

of our cross-country matrix (Table 5), where high-income non-English economies show a near-zero productivity response.

The same variable drives the experiment and the cross-country analysis. Extending the market-entry task from our four core languages to ten languages that span the resourcedness range, and regressing value retained on log Common Crawl share—the input to our macro Language Intensity measure—yields a positive, significant slope ($b = 0.07$, SE 0.03, $t = 2.4$; Figure 4(b)). We treat this regression as illustrative: it rests on a small, purposively chosen set of languages and a single binary-outcome task, and a comprehensive analysis matching our macro sample is left to future work. The direction is nonetheless unambiguous and sharpest at the low-resource extreme. The mechanism we estimate at the level of national productivity is thus the same one we see in individual business decisions: compatibility with training data, not income, infrastructure, or model origin, governs how much economic value AI delivers.

8.3 The Economic Cost: A Double Penalty

Because each decision carries a monetary stake, value lost translates into forfeited profit. On the production-planning decision, whose optimal monthly profit averages $7,613, a firm operating in English forfeits $824 (11%) by executing the AI's recommended plan; the *identical* decision in Yoruba forfeits $3,322 (44%)—about four times as much. Low-resource users face a second penalty on the cost side: because tokenizers fragment underrepresented languages, the same problem consumes two to three times as many input tokens in Yoruba as in English, and weaker models fail to return a usable decision 15–25% of the time in Yoruba versus near zero in English and Chinese. Lower-value decisions, at higher cost per query, more often with no usable output at all—a double penalty borne by exactly the languages least represented in training data (Table 9).

Table 9: The double penalty of low linguistic compatibility

	English	Chinese	Swahili	Yoruba
Forfeited profit per production decision (%)	11	10	22	44
Input-token cost (relative to English)	1.0×	1.1×	1.4×	∼3×
Unusable-output rate, weaker models (%)	∼0	∼0	6–15	15–25

These two penalties compound. A firm operating in Yoruba captures less than half the decision value of one operating in English *and* pays several times more per query to obtain it, so its effective cost per unit of realized decision value is many times higher—before counting the roughly one query in five that returns nothing usable at all. This is a channel distinct from the productivity effect our macro analysis estimates: even holding a decision's stakes fixed, the cost of *accessing* competent AI assistance rises as linguistic compatibility falls. The burden lands on exactly the low-resource-language populations our framework identifies as least represented in training corpora, widening the divide at a second margin—these users not only realize less value from each AI-assisted decision, they pay more to obtain it.

8.4 Sovereign Models Reach Parity in the Home Language

Our sovereign-AI counterfactual (Section 9.5) models a domestic model as raising a large economy's effective λ, and the experiment supports that premise directly. On Chinese-language decisions, the leading Chinese model (DeepSeek-V3) matches or exceeds GPT-4.1-mini—a capable, widely deployed Western model in the same low-cost tier—with value retained of 1.00 vs. 1.00 on production planning, 0.94 vs. 0.91 on shadow-price analysis, and 1.00 vs. 0.90 on value-of-information. The lesson is not that domestic models are superior but that they are *fully competitive* in the home language—precisely what a sovereign-AI path requires.

This also marks the boundary of that path. A competitive domestic model is feasible only where the home language is itself well represented in training data and the economy is large enough to fund the effort; under those conditions a country can capture AI's economic value on its own terms—the mechanism behind the leapfrogging dynamic of the preceding section. The populations left behind are those whose languages are too underrepresented to support either a competitive domestic model or adequate coverage in the global frontier—the lowest-resource end of our experiment, where value collapses regardless of model origin. For them the divide narrows only as the frontier's training data broadens to include their languages, the policy distinction we return to in Section 10.3.

9 Quantitative Analysis

The reduced-form estimates in Sections 5–7 establish the causal effect of linguistic compatibility on AI-driven productivity growth at current AI levels. But generative AI is advancing at an unprecedented pace—the capabilities of frontier models have transformed dramatically in just three years since ChatGPT's release—and the inequality patterns we document are only the beginning. As AI becomes more central to production, will the linguistic divide deepen? How large is the global output cost of monolingual training data? Can multilingual

AI development or sovereign AI strategies alter the trajectory? These questions lie beyond the reach of reduced-form estimation but can be addressed through the structural model developed in Section 2. We calibrate the model to match our empirical estimate, then conduct counterfactual experiments that quantify the stakes of policy inaction and the returns to alternative paths.

9.1 Calibration

The model's steady-state equation (4) contains one free parameter: the AI productivity parameter μ. All other parameters are either set externally or recovered from data.

External parameters. We set the discount rate $\rho = 0.02$ and the frontier growth rate $g_W = 0.015$, both standard in the growth literature. The elasticity of substitution $\sigma = 5$, implying per-intermediate profit $\pi = 1/\sigma = 0.2$, follows Acemoglu et al. (2026). The adoption cost Γ is a free normalization set to 1. The distance-to-frontier elasticity $\varepsilon = 1.1$ follows AAJ's baseline parameterization. We show in Appendix E that all counterfactual quantities are essentially insensitive to ε within the range $[0.8, 1.5]$.

Linguistic compatibility λ_i. For each country, λ_i equals the empirical Language Intensity variable normalized to $[0, 1]$ by dividing by the sample maximum. This maps Nigeria, Ghana, and the United Kingdom to $\lambda = 1.00$; China to $\lambda = 0.13$; Japan to $\lambda = 0.11$; and Korea to $\lambda = 0.02$.

Pre-AI absorptive capacity. The composite $\theta_i\beta_i$ is recovered from the pre-AI steady-state condition (5). We use each country's average labor productivity relative to the United States over the pre-ChatGPT period (2015–2021) as a proxy for $q_i^{*,\text{pre}}$ and invert the equation:

$$\theta_i\beta_i = \frac{\Gamma(\rho + g_W)}{\pi} \cdot \left(q_i^{*,\text{pre}}\right)^{\varepsilon}. \tag{11}$$

This ensures the model matches the pre-AI cross-country income distribution by construction, so that all counterfactual variation comes from the AI channel.

Calibrating μ. The model predicts long-run steady-state changes, while our regression estimates short-run annual growth differentials. We bridge the two through convergence dynamics. Near the balanced growth path, a country's annual growth acceleration following the μ shock is approximately:

$$\text{Annual growth acceleration}_i \;\approx\; r \times \left[(1 + \lambda_i \mu)^{1/\varepsilon} - 1\right], \tag{12}$$

where r is the annual rate at which a country's productivity adjusts toward its new post-AI steady state. We adopt $r = 0.10$ as our baseline, following the GMM panel estimate of Caselli et al. (1996), whose identification strategy—exploiting cross-country variation conditional on country fixed effects—most closely parallels our own.[4]

Given r and ε, we compute the model-predicted annual growth acceleration for each of the 169 economies at a candidate value of μ, regress these accelerations on standardized Language Intensity, and search for μ such that the regression slope equals our empirical estimate of 0.94 percentage points per standard deviation. This procedure yields $\boldsymbol{\mu = 0.359}$. The implied steady-state AI gains are 32.2 percent for fully English-compatible economies ($\lambda = 1$), 4.3 percent for China ($\lambda = 0.13$), 3.4 percent for Japan ($\lambda = 0.11$), and 0.5 percent for Korea ($\lambda = 0.02$). Table 10 summarizes the calibrated parameters.

What is and is not identified by the data. A useful way to interpret the calibration is to ask what cross-country variation in productivity growth pins down. The linearization of (12) reveals that the slope of growth accelerations on Language Intensity is governed by the

[4]The convergence literature reports a wide range of estimates depending on methodology and conditioning set. Mankiw et al. (1992) derive a theoretical Solow-style local rate of approximately 0.04 from $(1-\alpha)(n+g+\delta)$; Islam (1995) obtains around 0.05 from panel methods with country fixed effects; Caselli et al. (1996) obtain approximately 0.10 from system GMM; **?** argue that the higher GMM estimates partly reflect dynamic panel bias. Our baseline $r = 0.10$ sits at the upper end of this range, making the calibrated μ—and the implied counterfactual magnitudes—conservative relative to lower values of r.

Table 10: Calibrated Parameters

Parameter	Symbol	Value	Source
Discount rate	ρ	0.02	Standard
Frontier growth rate	g_W	0.015	Standard
Elasticity of substitution	σ	5	Acemoglu et al. (2026)
Adoption cost	Γ	1	Normalization
Distance-to-frontier elasticity	ε	1.1	Acemoglu et al. (2026)
Convergence rate	r	0.10	Caselli et al. (1996)
Linguistic compatibility	λ_i	Data	Language Intensity (normalized)
Pre-AI absorptive capacity	$\theta_i\beta_i$	Eq. (11)	Match pre-AI productivity
AI productivity	$\boldsymbol{\mu}$	**0.359**	**Match 0.94 pp estimate**

product $r\mu/\varepsilon$, not by r or μ separately. Cross-country DID therefore identifies this product to high precision but cannot distinguish between (r, μ) combinations that leave the product unchanged. Three implications follow. First, the choice of r shifts the calibrated μ in a predictable way and rescales country-level steady-state gains, dollar magnitudes, and changes in the inequality measures. Second, all comparative statics that depend only on the slope—ranking of country gains, sign of the inequality change, and ratios between counterfactual scenarios—are invariant to r. Third, ε is essentially absorbed into μ through the calibration and so plays a negligible role in counterfactual outcomes. Appendix E reports the full robustness of every quantitative claim in this section to $r \in [0.04, 0.15]$ and $\varepsilon \in [0.8, 1.5]$, confirming that the qualitative conclusions hold throughout, while the magnitudes scale with r in the manner just described.

9.2 Country Trajectories and Leapfrogging

We begin by examining how steady-state productivity evolves as AI becomes more central to production. For each country, we compute $q_i^*(\mu') = q_i^{*,\text{pre}} \times (1 + \lambda_i\mu')^{1/\varepsilon}$ across a range of AI importance levels $\mu' \in [0,\ 10\mu]$, where $\mu' = \mu$ corresponds to current AI levels and higher multiples represent scenarios in which AI capabilities and economic integration continue to advance.

Figure 5 presents absolute steady-state trajectories for four illustrative economies. Nigeria and the United Kingdom, both with $\lambda = 1.00$, accelerate in lockstep: their lines diverge sharply from those of Japan ($\lambda = 0.11$) and Korea ($\lambda = 0.02$), which barely rise. The divergence *widens* as AI importance grows—the linguistic divide is not a fixed gap but an expanding one.

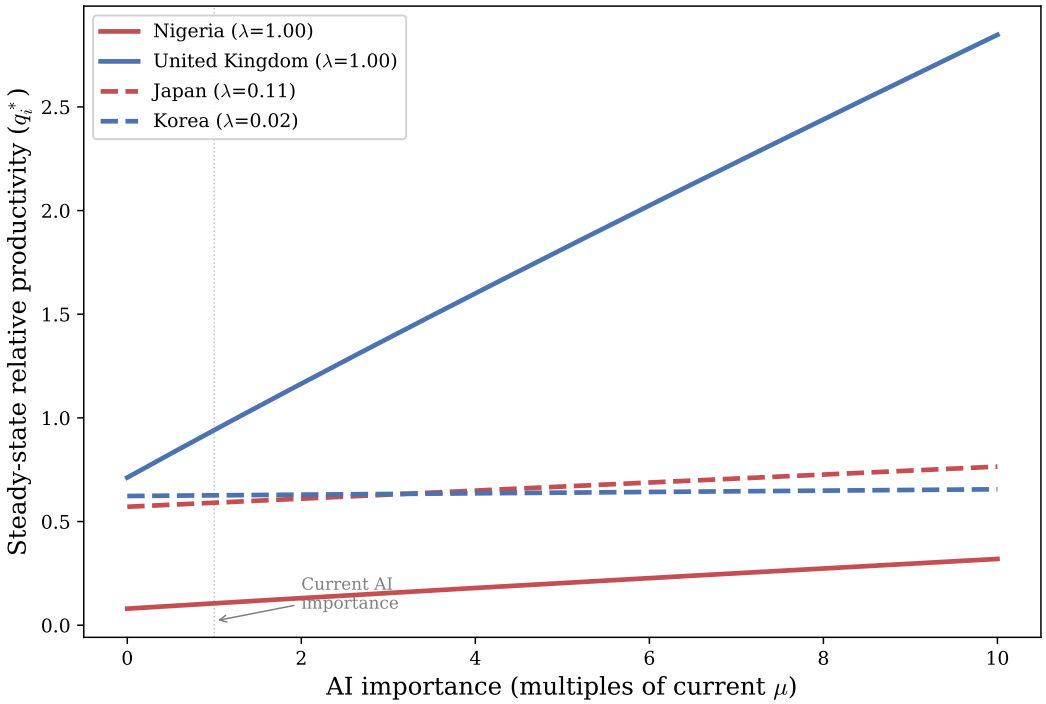


Figure 5: Steady-State Productivity Under Rising AI Importance

Notes: Each line plots $q_i^* = q_i^{*,\text{pre}} \times (1+\lambda_i \mu')^{1/\varepsilon}$ as a function of AI importance (multiples of current $\mu = 0.359$). Solid lines: high-λ economies (Nigeria, United Kingdom; $\lambda = 1.00$). Dashed lines: low-λ economies (Japan, $\lambda = 0.11$; Korea, $\lambda = 0.02$). $\varepsilon = 1.1$.

Figure 6 takes this further by plotting the productivity of several developing economies *relative to Japan*. When this ratio exceeds unity, leapfrogging has occurred. Ghana ($\lambda = 1.00$) approaches the crossover point as AI importance rises to roughly ten times current levels; Thailand ($\lambda = 0.01$) never approaches it. The direction of convergence is determined by λ; only the timeline depends on how rapidly μ grows.

These projections should be interpreted as illustrative of the model's qualitative predictions rather than as precise forecasts: at higher multiples of μ, the model is increasingly an

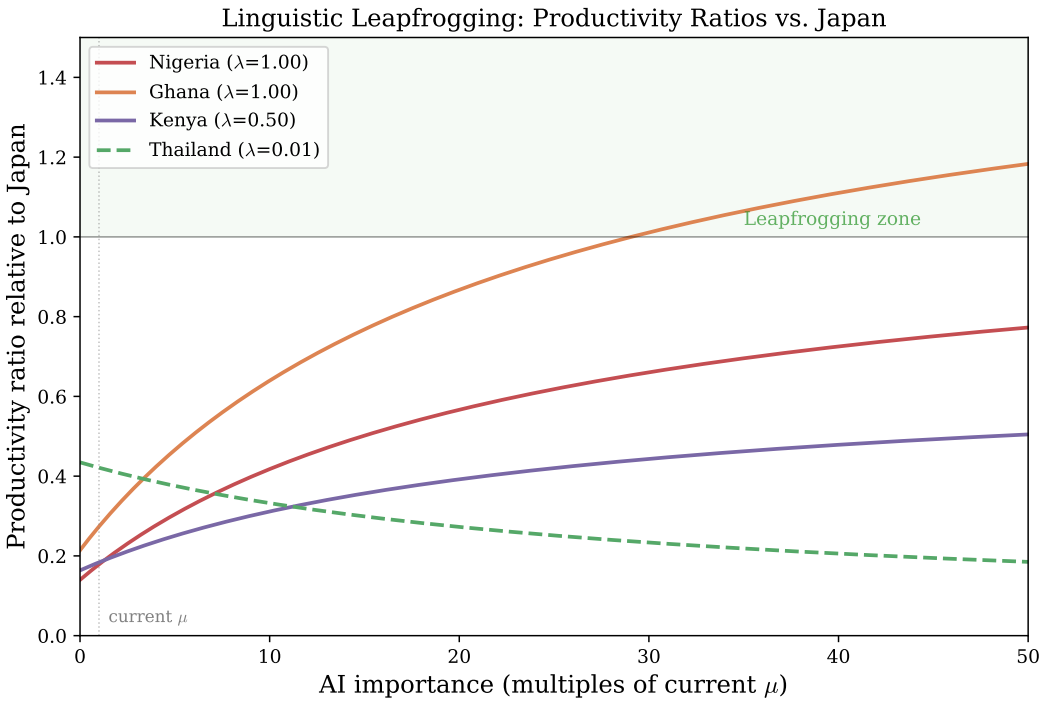


Figure 6: Linguistic Leapfrogging: Productivity Ratios Relative to Japan

Notes: Each line plots $q_P^*(\mu')/q_{\text{Japan}}^*(\mu')$ as a function of AI importance. A ratio exceeding unity indicates leapfrogging. Ghana and Nigeria ($\lambda = 1.00$) approach the crossover; Thailand ($\lambda = 0.01$) does not. $\varepsilon = 1.1$.

approximation, and structural changes—such as endogenous multilingual AI development—would alter the trajectory.

9.3 The Population Perspective: Who Is Left Behind?

Cross-country averages can obscure the human scale of AI-driven divergence. To translate the model's predictions into terms that reflect the distribution of gains across the world's population, we weight each country's AI productivity gain by its labor force.

Figure 7 plots the cumulative distribution of world workers against AI-induced productivity gains at current μ. The curve reveals a stark concentration of benefits. India (approximately 510 million workers, $\lambda \approx 0$) contributes a vertical jump at near-zero gains: roughly 22 percent of the world's labor force receives essentially no AI benefit.[5] Adding China (750 million workers, $\lambda = 0.13$) brings the cumulative share to 84 percent—meaning that at current AI levels, **84 percent of the world's workers gain less than 4.3 percent from AI**. At the opposite extreme, English-speaking economies account for only 5 percent of the global labor force but capture steady-state gains exceeding 30 percent.

The population-weighted perspective also reveals that standard inequality measures substantially understate the problem. The equal-weight variance of $\ln q_i^*$ increases by 0.4 percent after AI's arrival; the population-weighted variance increases by 5.0 percent—an order of magnitude larger. The population-weighted Gini coefficient of cross-country productivity rises from 0.409 to 0.431. In dollar terms, we estimate that if below-average-λ economies could achieve average linguistic compatibility, the resulting output gain would be approximately **$3.6 trillion per year** at current AI levels—a figure that grows as AI becomes more central to production.

[5]India's near-zero λ reflects the fact that our Language Intensity measure is based on mother-tongue data from the CIA World Factbook. While India has a substantial English-speaking population (approximately 10 percent as a second language), the majority of the labor force works in languages with minimal representation in AI training corpora. This is a known limitation of the measure and, if anything, makes our estimates conservative for India.

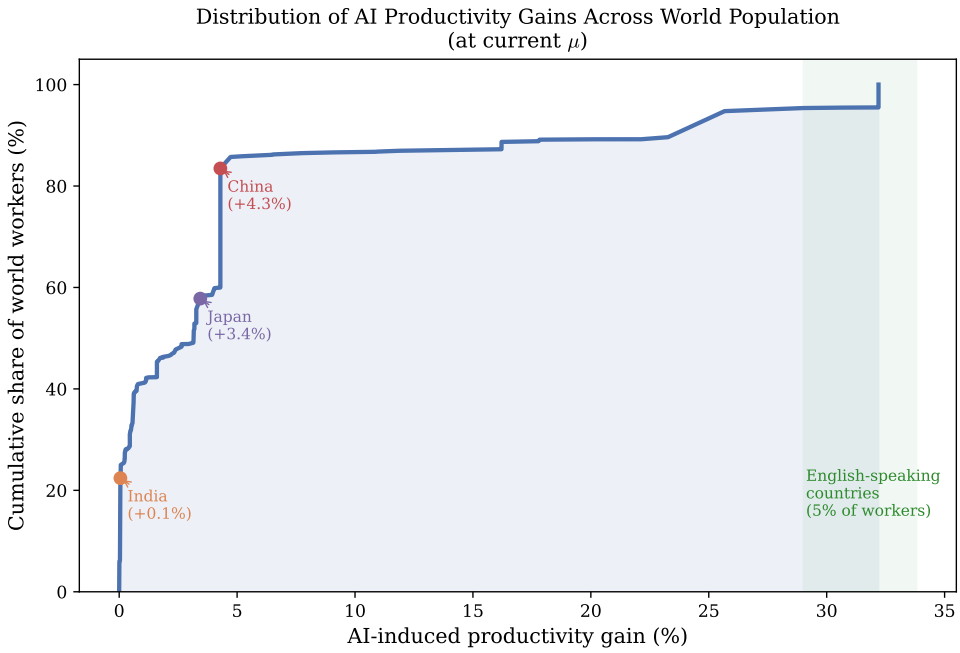


Figure 7: Distribution of AI Productivity Gains Across World Population

Notes: Horizontal axis: steady-state AI productivity gain $[(1+\lambda_i\mu)^{1/\varepsilon}-1]\times 100$ at current $\mu = 0.359$. Vertical axis: cumulative share of world workers (employment-weighted). Labor force approximated as GDP (PPP) divided by output per worker. India ($\lambda \approx 0$, 510M workers) contributes the jump near zero; China ($\lambda = 0.13$, 750M workers) brings the cumulative share to 84%.

9.4 Multilingual AI: The Growth–Inequality Tradeoff

A central policy question is whether the inequality cost of AI-driven growth is inherent or contingent on the linguistic composition of training data. We address this through two complementary analyses.

Figure 8 presents kernel density estimates of the world productivity distribution under rising AI importance. The left panel shows that as μ increases, the distribution's right tail stretches progressively—high-λ economies pull ahead while the left tail remains largely unchanged. The right panel compares the status quo distribution at $10\times\mu$ with a multilingual counterfactual in which each country's λ_i is compressed toward the cross-country mean: $\lambda_i^{\text{cf}} = 0.5\lambda_i + 0.5\bar{\lambda}$.[6] The multilingual distribution is substantially more compressed.

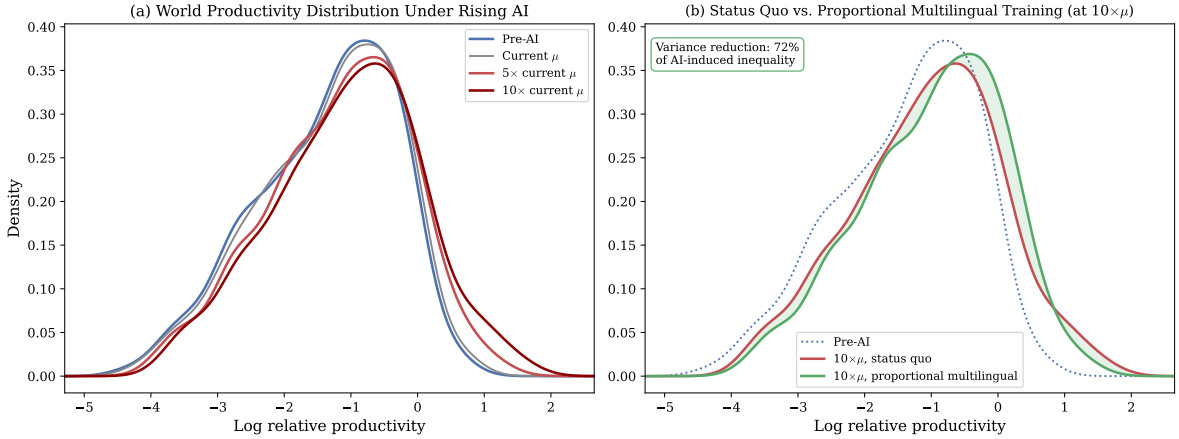


Figure 8: World Productivity Distribution Under Rising AI

Notes: Kernel density estimates of $\ln q_i^*$ across 169 economies (bandwidth = 0.3). Left panel: distributions under pre-AI, current μ, $5 \times \mu$, and $10 \times \mu$. Right panel: status quo versus proportional multilingual counterfactual ($\lambda_i^{\text{cf}} = 0.5\lambda_i + 0.5\bar{\lambda}$) at $10 \times \mu$. Equal country weights.

Figure 9 makes the policy tradeoff explicit. For each value of μ, we plot global output growth (horizontal axis) against the increase in population-weighted $\text{Var}(\ln q_i^*)$ relative to the pre-AI baseline (vertical axis). Under the status quo, each percentage point of AI-driven global growth is accompanied by a 0.39 percentage-point increase in cross-country

[6] We adopt this compression rather than setting $\lambda_i = 1$ for all economies, which would be unrealistically extreme. The counterfactual should be interpreted as a proportional increase in multilingual representation in training data, not complete linguistic equalization. The specific parameter (0.5) is illustrative.

inequality. Under multilingual AI, the same growth comes at roughly half the inequality cost (0.21 percentage points)—at $10 \times \mu$, multilingual training reduces the inequality cost of AI-driven growth by 59 percent. The shaded area between the two curves represents the inequality savings from multilingual training: a Pareto improvement that does not reduce gains for English-speaking economies but extends them to others.

This analysis reframes multilingual AI development from a purely equity-motivated intervention to an efficiency argument: it is not that multilingual AI sacrifices growth for fairness, but that it achieves the same growth at a lower inequality cost. Training data composition is a design choice, and different choices yield quantitatively different tradeoffs.

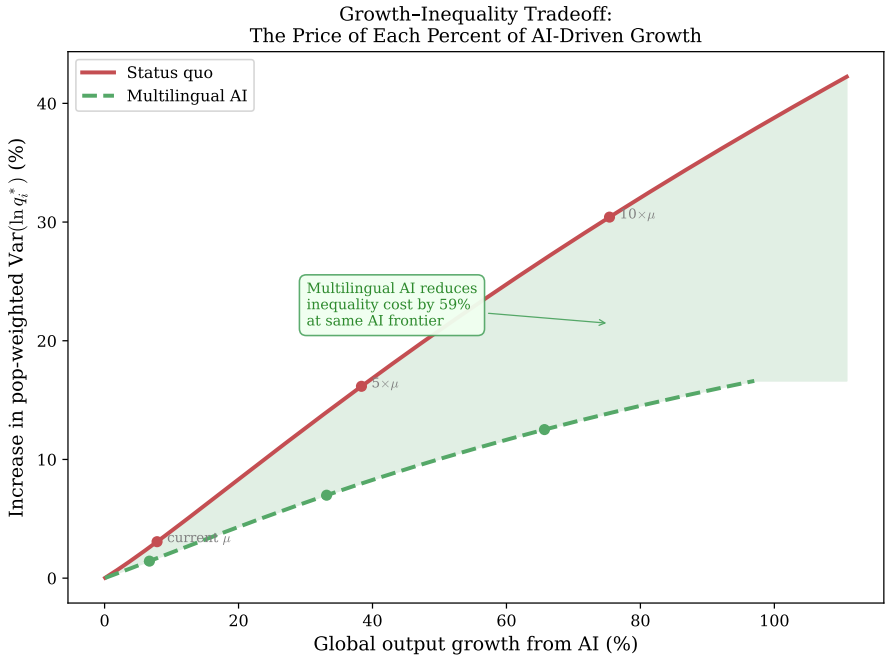


Figure 9: Growth–Inequality Tradeoff: The Price of AI-Driven Growth

Notes: Each point on a curve corresponds to a different AI importance level (μ' ranging from 0.01μ to 15μ). Horizontal axis: global output growth from AI (%, population-weighted). Vertical axis: increase in population-weighted $\text{Var}(\ln q_i^*)$ relative to pre-AI baseline (%). Shaded area: inequality savings from multilingual training. Dots mark current μ, $5 \times \mu$, and $10 \times \mu$.

9.5 Sovereign AI: A Complementary Path

A complementary response to the linguistic constraint we identify is sovereign AI: individual economies developing AI systems optimized for their domestic languages. China's rapid development of large language models—including DeepSeek and Qwen, trained substantially on Chinese-language corpora—provides a natural case study of how a large economy can expand its own linguistic compatibility through domestic investment. We model this scenario by raising China's effective λ from 0.13 to 0.70 while holding other economies' parameters fixed.

Figure 10 presents population-weighted kernel density estimates of the world productivity distribution under this scenario, at both $5 \times \mu$ and $10 \times \mu$. At $5 \times \mu$, China shifts rightward in the distribution while other economies remain near their status-quo positions. At $10 \times \mu$, the distribution becomes *bimodal*: one peak corresponds to the mass of low-λ economies, and a second peak emerges as China joins the high-λ cluster alongside English-speaking economies. The population-weighted Gini coefficient falls relative to the status quo, reflecting that China—representing 23 percent of the world's labor force—moves up in the distribution. The structural pattern shifts from a unipolar one ("English versus the rest") to a bimodal one ("English and Chinese versus the rest").

This analysis carries two implications. First, sovereign AI is an effective response for economies with the scale and resources to develop domestic large language models: it directly raises a country's effective λ and delivers steady-state gains comparable to those of high-λ economies. Second, because the gains accrue specifically to the investing country, sovereign AI and multilingual AI address different scopes of the problem—the former works at the national scale, the latter at the global scale. The two paths are complements rather than substitutes: sovereign AI expands compatibility for individual large economies, while multilingual training expands it for the many smaller economies that lack the resources to develop independent domestic models.

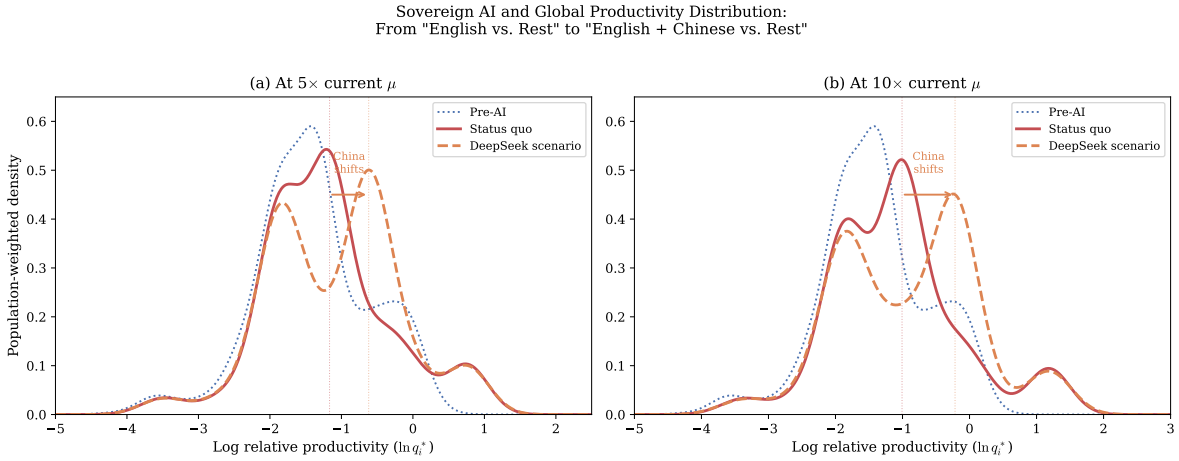


Figure 10: Sovereign AI and Global Productivity Distribution

Notes: Population-weighted kernel density estimates of $\ln q_i^*$. DeepSeek scenario: China's effective λ raised from 0.13 to 0.70; all other economies unchanged. Left panel: $5 \times \mu$. Right panel: $10 \times \mu$. The bimodal structure at $10 \times \mu$ reflects China joining the high-λ cluster alongside English-speaking economies.

9.6 Summary

The calibrated model produces several quantitative predictions. At current AI levels, 84 percent of the world's workers gain less than 4.3 percent from AI, while 5 percent of the labor force captures gains exceeding 30 percent. The population-weighted Gini coefficient rises from 0.409 to 0.431, and the English-language bias in training data corresponds to an estimated $3.6 trillion in forfeited annual output. As AI importance grows, the divergence widens through the model's steady-state structure rather than narrowing. Multilingual AI development reduces the inequality cost of AI-driven growth by roughly 59 percent at the global level, while sovereign AI raises compatibility for individual large economies—two complementary paths we return to in the discussion that follows.[7]

[7]The qualitative conclusions of this section—the direction of leapfrogging, the inequality-reducing effect of multilingual AI, the bimodal restructuring under sovereign AI—are invariant to the choice of r. Quantitative magnitudes (calibrated μ, country-specific gains, dollar losses) scale with r. As shown in Appendix E, varying r across the literature range $[0.04, 0.15]$ leaves all qualitative conclusions intact while changing the dollar loss estimate from $2.4 to $9.2 trillion. Our baseline of $3.6 trillion sits at the conservative end of this range.

10 Discussion

10.1 Theoretical Implications

Our theoretical model and quantitative analysis yield three insights that extend beyond the reduced-form findings.

Traditional inequality measures understate the problem. The equal-weight cross-country variance of $\ln q_i^*$ increases by 0.4 percent following AI's arrival—a seemingly modest figure that has featured in prior discussions of AI and global inequality. But this statistic treats Nigeria and China as equally weighted observations. When we weight by labor force, the variance increase is 5.0 percent—an order of magnitude larger. The discrepancy arises because the world's two most populous economies, China (750 million workers, $\lambda = 0.13$) and India (500 million workers, $\lambda \approx 0$), together account for roughly 40 percent of the global labor force and both fall in the left tail of the linguistic compatibility distribution. Standard cross-country metrics miss the human scale of AI-driven divergence: at current AI levels, 84 percent of the world's workers gain less than 4.3 percent, while 5 percent capture gains exceeding 30 percent. Standard cross-country measures and the population-weighted view answer different questions: the former tracks divergence among countries, the latter among workers. Both are informative; population weighting reveals a dimension that country-level statistics, by construction, do not.

Sovereign AI changes the structure of inequality, not just the level. China's rapid development of domestic large language models—DeepSeek, Qwen, and others trained substantially on Chinese-language corpora—represents a concrete policy response to the linguistic constraint our model identifies. Our counterfactual analysis shows that if China raises its effective λ from 0.13 to 0.70 through sovereign AI development, its steady-state productivity gain rises from 4.3 to roughly 23 percent—a substantial expansion of compatibility for nearly a quarter of the world's labor force. The structural consequence is a shift from a unimodal pattern ("English versus the rest") to a bimodal one ("English and Chinese versus

the rest"). This is a qualitatively new form of global stratification absent from existing theoretical frameworks, and it implies that the relevant policy question is not simply whether AI increases inequality, but how its architecture is reshaped by which languages AI systems are built to serve.

Growth and inequality are not inevitably linked. A common implicit assumption in discussions of AI and inequality is that faster AI-driven growth necessarily comes at the cost of greater cross-country divergence. Our growth–inequality tradeoff analysis shows this is contingent on training data composition. Under the status quo, each 1 percent of global output growth from AI is accompanied by a 0.39 percent increase in population-weighted inequality. Under proportional multilingual training, the same growth comes at roughly half the inequality cost. Multilingual AI development is therefore not merely an equity intervention—it is an efficiency improvement that makes AI-driven growth "cheaper" in terms of its distributional consequences. This reframing matters for policy: it shifts the case for multilingual AI from a normative argument about fairness to a positive argument about the global returns to a public investment in training data diversity.

10.2 Policy Implications

Our findings reveal a fundamental tension: the linguistic bias embedded in AI training data is a design choice with large distributional consequences, and different choices yield quantitatively different outcomes. We organize policy responses along three time horizons.

10.2.1 The Short-Term Dilemma: English Education

For developing economies, investing in English-language education offers a pragmatic short-term response. Our evidence shows that English proficiency enables poor economies to capture AI productivity gains comparable to those of wealthy Anglophone nations (Table 5). From a pure economic efficiency standpoint, population-level English investment may yield high returns in the AI era, particularly when combined with digital infrastructure

(Section 7.2).

Yet this path carries troubling implications. Our calibration shows that even a substantial investment in English education—raising λ by 0.1—yields only 2–3 percent in steady-state productivity gains, a modest return relative to the supply-side alternative of multilingual AI development. More fundamentally, requiring the world to learn English to access AI's benefits constitutes a form of linguistic imperialism: it privileges a single language—one tied to specific colonial and economic histories—as the gateway to technological progress. As we argued in Section 2.1, language shapes thought itself; one's native language is not merely a different label for universal concepts but a distinct cognitive framework. Forcing populations to adopt English for AI access imposes cognitive costs and undermines the linguistic diversity that international frameworks such as the Universal Declaration of Linguistic Rights seek to protect.

10.2.2 The Medium-Term Imperative: Multilingual AI Development

The quantitative analysis provides concrete grounding for what has previously been a normative aspiration. At current AI levels, the English-language bias in training data costs the world an estimated $3.6 trillion in forfeited annual output—a figure that grows as AI becomes more central to production. Multilingual training would cut the inequality cost of AI-driven growth by 59 percent, achieving the same global output gains at substantially lower distributional cost.

These are not merely theoretical projections. NVIDIA and the Singaporean government jointly released Nemotron-Personas-Singapore in January 2026, a synthetic dataset supporting sovereign AI development across Singapore's four official languages—an early signal that multilingual AI development is transitioning from aspiration to policy reality. Our analysis provides the economic case for scaling such efforts: multilingual AI is a Pareto improvement that does not reduce gains for English-speaking economies but extends them to others.

For international organizations and donors, funding multilingual AI development for un-

derrepresented languages warrants recognition as a global public good. Individual developing economies lack the resources and incentives to unilaterally develop AI systems for their languages, particularly when populations number in the millions rather than hundreds of millions. Mechanisms could include multilateral funding for open-source multilingual models, prizes for low-resource language AI performance, or technical assistance for economies developing localized AI systems.

10.2.3 The Longer View: Linguistic Rights in the AI Era

Beyond immediate policy responses, our findings raise fundamental questions about linguistic rights in an AI-driven economy. If productivity increasingly depends on AI tools, and those tools perform differentially across languages, then linguistic compatibility becomes an economic necessity rather than a cultural preference. Populations speaking underrepresented languages face a choice that no prior technology has imposed: abandon their native tongues for economic participation, or accept productivity penalties for linguistic preservation. This is a form of linguistic imperialism that our quantitative findings make visible—not as an abstract concern but as a measurable cost borne disproportionately by already-disadvantaged populations.

Sovereign AI development addresses this dilemma at the national scale. China's development of DeepSeek and Qwen demonstrates that large economies with sufficient resources can directly expand their own linguistic compatibility, capturing AI's productivity gains on their own terms rather than waiting for global training data to broaden. Yet many smaller economies—particularly those whose languages are spoken by populations too small to support domestic frontier-model development—cannot follow the same path. For these economies, the binding constraint can only be relaxed by broader multilingual AI development at the global level, which expands the linguistic reach of training data for the many languages that individual sovereign AI strategies cannot economically cover.

The pattern we document is not an inevitable consequence of AI but a consequence of

choices about how AI systems are built. Training data composition, model architecture, and resource allocation are design decisions. Our evidence shows that these decisions carry distributional consequences of the first order—and that different decisions can yield fundamentally different outcomes for global inequality.

10.3 Limitations and Future Research

Several limitations warrant acknowledgment.

Our analysis relies on macro-level productivity data, with all the identification challenges that cross-country designs entail. We cannot directly observe how individuals and firms use AI, or trace the precise pathways through which linguistic barriers reduce effectiveness. Firm-level studies, patent-based analyses, and task-level experiments—some of which we are pursuing in companion work—offer sharper identification and will complement the patterns we document here.

Yet the trade-off is itself worth reflecting on. The cleanliness of causal identification is closely tied to data availability, and data availability maps onto the very axis of global inequality this paper studies. The economies whose languages occupy effectively zero percent of Common Crawl are the same economies that lack firm registries, linked employer-employee data, and representative digital adoption surveys. Research designs that prioritize identification rigor tend, by construction, to concentrate on data-rich economies and to exclude precisely those that our framework identifies as most vulnerable to linguistic exclusion. In this sense, methodological constraints replicate the inequality that our substantive findings describe.

Posing this question at the cross-country level is therefore not a concession to data limitations but a deliberate choice about what the AI era's distributional question requires. If we study AI's impact only where micro data are abundant, the economies most likely to be left behind will remain invisible in the empirical record—just as they are in the training corpora themselves. Within this framework, ILO labor productivity growth rates represent the best

available dependent variable, and the 168-economy sample represents the appropriate unit of observation for a question about global technology diffusion.

The three-year post-ChatGPT window (2022–2024) is short by the standards of technology diffusion research, though unavoidably so: generative AI at scale is itself only three years old. That we detect significant and persistent productivity divergence within this window suggests the linguistic stratification we document is a structural feature of how this technology operates rather than a transitional artifact. Whether translation technologies or multilingual models will eventually narrow these gaps remains an open question that only time and continued monitoring can answer.

The quantitative analysis introduces additional limitations specific to calibration. The convergence rate r is set externally rather than estimated, and the specific value of μ depends on this choice: higher r implies lower μ and correspondingly smaller steady-state gains, while lower r implies the reverse. The multilingual counterfactual compresses λ toward its mean using a specific functional form; results depend on this parameterization and should be interpreted as illustrative rather than as precise policy targets. At high multiples of current μ, the model's linear-in-λ adoption structure is increasingly an approximation: structural changes in the economy, endogenous multilingual AI development, and shifts in the production function itself would alter trajectories in ways the static model cannot capture. We interpret projections beyond $5 \times \mu$ as stress tests indicating the direction and broad magnitude of effects rather than point estimates.

Our Language Intensity measure is based on mother-tongue data from the CIA World Factbook, which captures first-language speakers but not second-language proficiency. For economies with substantial bilingual populations—most notably India, where approximately 10 percent of the population speaks English as a second language—this likely understates effective linguistic compatibility with AI training data. If anything, this makes our baseline estimates conservative: the true relationship between linguistic compatibility and AI productivity gains may be stronger than what we report.

Sectoral and gender analyses presented in Appendix B are explicitly exploratory, offered as preliminary evidence of where linguistic compatibility effects manifest within economies rather than as settled findings. These patterns point toward productive directions for future work with higher-quality sectoral and individual-level data.

11 Conclusion

This paper provides the first causal evidence that artificial intelligence amplifies global inequality through linguistic compatibility with training data. Exploiting ChatGPT's November 2022 release as a natural experiment across 169 economies, we find that a one-standard-deviation increase in linguistic compatibility raises productivity growth by 0.94 percentage points—a 58 percent acceleration over the baseline rate. This effect is robust to controls for institutions, economic development, digital infrastructure, and AI exposure intensity.

The distributional pattern overturns conventional expectations. Low-income English-speaking economies—Ghana, Nigeria, Kenya—experience productivity gains comparable to wealthy Anglophone nations, while high-income non-English economies—Japan, South Korea, Germany—show no significant effects despite superior resources. Nigeria, with GDP per capita of approximately $2,447, captures AI productivity gains; Japan, at $33,800, does not. This "linguistic leapfrogging" challenges conventional technology diffusion theory and reveals that language compatibility functions as a necessary condition for AI effectiveness that wealth and institutions cannot substitute for.

We demonstrate this through three complementary analyses. Colonial history provides quasi-exogenous variation: among former colonies, only English-speaking ones benefit, mirroring training data composition (44.3 percent English, less than 5 percent French or Spanish in Common Crawl). Results reflect differential effectiveness rather than differential exposure, confirmed by controls for AI awareness and uptake. Within-colonial-sample analysis shows poor English-speaking economies outperform wealthy non-English ones, isolating language

from institutional and historical confounds.

Theoretically, we extend the technology adoption framework of Acemoglu et al. (2026) by introducing linguistic compatibility as a new determinant of AI-era productivity. The model shows that language operates as a necessary condition—qualitatively stronger than standard inappropriate technology, where mismatch reduces but does not eliminate gains. When linguistic compatibility is zero, the AI channel is shut off entirely, regardless of a country's absorptive capacity or institutional quality. The model also delivers an explicit leapfrogging threshold: the level of AI importance beyond which poor English-speaking economies overtake wealthy non-English ones. These results bridge economics and computer science by providing macro-level validation that laboratory performance gaps across languages (Lai et al., 2023; Ahuja et al., 2023) aggregate to national productivity divergence. A controlled experiment of our own makes this mechanism concrete: across languages, the economic value generative AI delivers on common business decisions, and the cost of obtaining it, track linguistic compatibility—with the steepest shortfalls in the least-represented languages.

Calibrating the model to our empirical estimate and projecting forward, the quantitative analysis reveals the full scale of the problem. At current AI levels, 84 percent of the world's workers gain less than 4.3 percent from AI, while 5 percent of the labor force captures gains exceeding 30 percent. The English-language bias in training data costs an estimated $3.6 trillion in forfeited annual output. As AI advances, the divergence widens rather than self-corrects. Sovereign AI development—as exemplified by China's DeepSeek—enables individual large economies to expand their own linguistic compatibility, producing a bimodal rather than unipolar distribution of AI-driven gains.

For policy, our findings reveal a fundamental tension. Developing economies could invest in English education to capture AI gains, but this constitutes a form of linguistic imperialism—privileging one language as the gateway to technological progress and undermining the cognitive and cultural role of native tongues. The sustainable solution is multilingual AI development: our calibration shows it would cut the inequality cost of AI-

driven growth by 59 percent, a Pareto improvement that extends benefits without reducing them for any country. The case for multilingual AI is therefore not only one of equity but also of efficiency: training data composition shapes who benefits from the technology, and different compositions yield quantitatively different outcomes. Multilingual AI requires coordinated investment from international organizations, AI developers, and governments who must recognize that linguistic inclusion is a precondition for equitable technological progress.

In the AI era, linguistic compatibility has become a binding constraint on productivity growth, creating unexpected winners and losers based on language rather than wealth. Nigeria leapfrogs Japan not despite its disadvantages, but because AI—for once—does not require the capital accumulation or institutional development that previous technologies demanded. The pattern we document is not inevitable but a consequence of choices about how AI systems are built. Different choices can yield different outcomes.

References

Acemoglu, Daron, "The simple macroeconomics of AI," *Economic Policy*, 2025, *40* (121), 13–58.

_ , Simon Johnson, and James A Robinson, "The colonial origins of comparative development: An empirical investigation," *American economic review*, 2001, *91* (5), 1369–1401.

_ , Ufuk Akcigit, and Simon Johnson, "Technology and Economic Development," Working Paper, National Bureau of Economic Research 2026.

Aghion, Philippe and Simon Bunel, "AI and Growth: Where Do We Stand?," Policy Note, EmergingTech Economic Research Network, Federal Reserve Bank of San Francisco June 2024.

Agrawal, Ajay, Joshua S. Gans, and Avi Goldfarb, "Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, Chicago, IL: University of Chicago Press, 2019, pp. 89–110.

Ahuja, Kabir, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Mohamed Ahmed, Kalika Bali, and Sunayana Sitaram, "MEGA: Multilingual Evaluation of Generative AI," in "Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing" 2023, pp. 4232–4267.

Bang, Yejin, Samuel Cahyawijaya, Nayeon Lee, Wenliang Dai, Dan Su, Bryan Wilie, Holy Lovenia, Ziwei Ji, Tiezheng Yu, Willy Chung, Quyet V. Do, Yan Xu, and Pascale Fung, "A Multitask, Multilingual, Multimodal Evaluation of ChatGPT on Reasoning, Hallucination, and Interactivity," in "Proceedings of the 13th International

Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics" 2023, pp. 675–718.

Bender, Emily M, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?," in "Proceedings of the 2021 ACM conference on fairness, accountability, and transparency" 2021, pp. 610–623.

Bick, Alexander, Adam Blandin, and David J Deming, "The rapid adoption of generative AI," *Management Science*, 2026.

Brown, Tom B, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell et al., "Language models are few-shot learners," in "Proceedings of the 34th International Conference on Neural Information Processing Systems" ACM 2020, pp. 1877–1901.

Brynjolfsson, Erik, Danielle Li, and Lindsey Raymond, "Generative AI at work," *The Quarterly Journal of Economics*, 2025, *140* (2), 889–942.

Caselli, Francesco, Gerardo Esquivel, and Fernando Lefort, "Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics," *Journal of Economic Growth*, 1996, *1* (3), 363–389.

Chaar, Tania, Francesco Filippucci, Cecilia Jona-Lasinio, and Giuseppe Nicoletti, "AI and the Global Productivity Divide: Fuel for the Fast or a Lift for the Laggards?," OECD Artificial Intelligence Papers 51, OECD Publishing December 2025.

Comin, Diego and Bart Hobijn, "An exploration of technology diffusion," *American economic review*, 2010, *100* (5), 2031–2059.

— **and Martí Mestieri**, "If technology has arrived everywhere, why has income diverged?," *American Economic Journal: Macroeconomics*, 2018, *10* (3), 137–178.

Council of Economic Advisers, "Artificial Intelligence and the Great Divergence," Technical Report, The White House January 2026.

Cruces, Guillermo, Diego Fernández Meijide, Sebastian Galiani, Ramiro H. Gálvez, and María Lombardi, "Does Generative AI Narrow Education-Based Productivity Gaps? Evidence from a Randomized Experiment," Working Paper, National Bureau of Economic Research 2026.

Dell'Acqua, Fabrizio, Edward McFowland III, Ethan R Mollick, Hila Lifshitz-Assaf, Katherine Kellogg, Saran Rajendran, Lisa Krayer, François Candelon, and Karim R Lakhani, "Navigating the jagged technological frontier: Field experimental evidence of the effects of AI on knowledge worker productivity and quality," *Harvard business school technology & operations mgt. Unit working paper*, 2023, (24-013).

Dodge, Jesse, Maarten Sap, Ana Marasović, William Agnew, Gabriel Ilharco, Dirk Groeneveld, Margaret Mitchell, and Matt Gardner, "Documenting Large Webtext Corpora: A Case Study on the Colossal Clean Crawled Corpus," in "Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing" Association for Computational Linguistics Online and Punta Cana, Dominican Republic 2021, pp. 1286–1305.

Filippucci, Francesco, Peter Gal, Katharina Laengle, and Matthias Schief, "Macroeconomic Productivity Gains from Artificial Intelligence in G7 Economies," OECD Artificial Intelligence Papers 41, OECD Publishing June 2025.

Gans, Joshua S., "Artificial Jagged Intelligence: When AI Benchmarks Misstate Deployment Value," Working Paper 34712, National Bureau of Economic Research 2026.

Gerschenkron, Alexander, *Economic Backwardness in Historical Perspective*, Cambridge, MA: Harvard University Press, 1962.

Goldfarb, Avi and Daniel Trefler, "Artificial Intelligence and International Trade," in Ajay Agrawal, Joshua Gans, and Avi Goldfarb, eds., *The Economics of Artificial Intelligence: An Agenda*, Chicago, IL: University of Chicago Press, 2019, pp. 463–492.

Humlum, Anders and Emilie Vestergaard, "The Unequal Adoption of ChatGPT Exacerbates Existing Inequalities among Workers," *Proceedings of the National Academy of Sciences*, 2025, *122* (1), e2414972121.

Islam, Nazrul, "Growth Empirics: A Panel Data Approach," *Quarterly Journal of Economics*, 1995, *110* (4), 1127–1170.

Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury, "The state and fate of linguistic diversity and inclusion in the NLP world," in "Proceedings of the 58th annual meeting of the association for computational linguistics" 2020, pp. 6282–6293.

Lai, Viet Dac, Nghia Ngo, Amir Pouran Ben Veyseh, Hieu Man, Franck Dernoncourt, Trung Bui, and Thien Huu Nguyen, "ChatGPT beyond english: Towards a comprehensive evaluation of large language models in multilingual learning," in "Findings of the association for computational linguistics: EMNLP 2023" 2023, pp. 13171–13189.

Mankiw, N. Gregory, David Romer, and David N. Weil, "A Contribution to the Empirics of Economic Growth," *The Quarterly Journal of Economics*, 1992, *107* (2), 407–437.

Melitz, Jacques and Farid Toubal, "Native language, spoken language, translation and trade," *Journal of International Economics*, 2014, *93* (2), 351–363.

Moscona, Jacob and Karthik A. Sastry, "Inappropriate Technology: Evidence from Global Agriculture," Working Paper, National Bureau of Economic Research 2025.

Noy, Shakked and Whitney Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, 2023, *381* (6654), 187–192.

Zhang, Xuan, Wenxuan Zhang, Anxu Wang, See-Kiong Ng, and Yang Deng, "MASim: Multilingual Agent-Based Simulation for Social Science," *arXiv preprint arXiv:2512.07195*, 2025.

A Appendix: Robustness to Inclusion of 2020

The following tables and figure replicate the main results using the full sample including 2020. The baseline coefficient remains stable, though pre-trend tests are less clean due to COVID-19's heterogeneous impacts across economies, as discussed in Section 3.

Table A1: Baseline DID Results

	Productivity Growth					
	Lang. Intensity		Primary Lang.		English Only	
	(1)	(2)	(3)	(4)	(5)	(6)
Lang. Int. × Post	0.874*	0.623**				
	(0.477)	(0.260)				
Primary × Post			0.072*	0.050**		
			(0.039)	(0.020)		
English × Post					0.065*	0.042**
					(0.038)	(0.021)
ln(GDP pc)		13.921***		13.850***		13.885***
		(4.023)		(4.013)		(4.066)
Internet		-3.891		-3.952		-4.015
		(2.675)		(2.666)		(2.686)
Observations	1,678	1,678	1,678	1,678	1,678	1,678
Adjusted R^2	0.210	0.257	0.211	0.257	0.209	0.256
Country FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Dependent variable: annual labor productivity growth (pp). Lang. Intensity: all languages ≥5% weighted by training data. Primary Lang. Only: primary language only. English Only: English share only. Post=1 for 2022+. Controls: ln(GDP per capita), Internet penetration. SE clustered at country level. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Appendix: Sectoral Heterogeneity and Labor Market Restructuring

The Language Intensity treatment constructed in this paper provides a lens through which to examine whether AI has generated measurable impacts on outcomes beyond aggregate labor

Table A2: Robustness: Controlling for AI Adoption and Structural Factors

	(1) g_outputpw	(2) g_outputpw
Language Intensity × Post	0.623**	0.629**
	(0.260)	(0.273)
AI Search Intensity (Std)		-0.146
		(0.153)
Log GDP per capita	13.921***	14.360***
	(4.023)	(4.027)
Internet penetration (%)	-3.891	-4.518
	(2.675)	(2.751)
Observations	1,678	1,598
Countries	169	161
Adjusted R^2	0.257	0.259
AI Control	No	Yes
full_controls	No	No

Notes: Dependent variable is annual labor productivity growth rate (percentage points).
AI Search Intensity measured using Google Trends data from Glimpse API, standardized.
Column 1 shows baseline specification. Column 2 adds AI adoption control.
All specifications include country and year fixed effects.
Standard errors clustered at country level. *** p<0.01, ** p<0.05, * p<0.1.

Table A3: Colonial History and Language Effects

	(1) British	(2) All Colonies	(3) By Type	(4) Brit vs Non-Brit
Language Intensity × Post	0.970**	0.903***		
	(0.404)	(0.337)		
Lang × Post × British Colony			0.977***	0.963**
			(0.370)	(0.365)
Lang × Post × French Colony			0.378	
			(1.111)	
Lang × Post × Spanish Colony			0.515	
			(1.794)	
Lang × Post × Non-British Colony				0.531
				(0.983)
Log GDP per capita	19.684***	17.166***	17.102***	17.121***
	(3.542)	(4.128)	(4.149)	(4.140)
Internet penetration (%)	1.141	1.539	1.414	1.462
	(6.129)	(3.935)	(3.984)	(3.930)
Observations	298	749	749	749
Countries	30	76	76	76
Adjusted R^2	0.540	0.373	0.371	0.372
Sample	British only	All colonies	All colonies	All colonies

Notes: Dependent variable is annual labor productivity growth rate (percentage points).

Column 1 restricts sample to former British colonies to control for institutional similarities.

Column 2 includes all former colonies (British, French, Spanish).

Column 3 separately estimates effects by colonial type.

Column 4 compares British vs non-British colonies .

All specifications include country and year fixed effects.

Standard errors clustered at country level. *** p<0.01, ** p<0.05, * p<0.1.

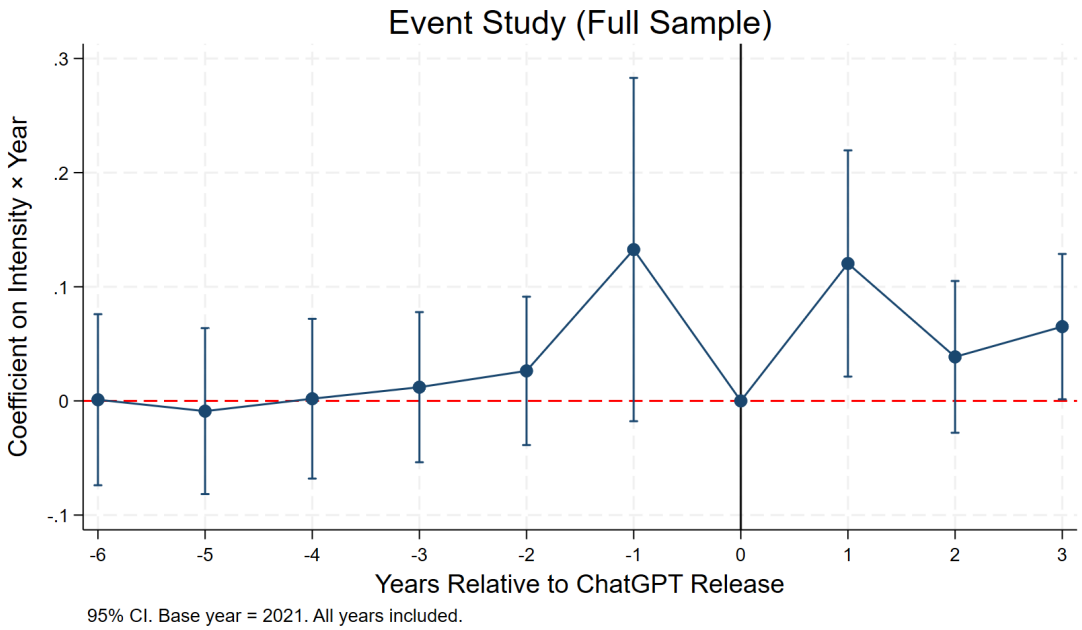


Figure A1: Event Study

productivity. We exploit this to offer exploratory evidence on two additional dimensions: sectoral heterogeneity and gender labor market patterns. We emphasize that these analyses are preliminary and intended to indicate productive directions for future research rather than to establish settled findings.

B.1 Sectoral Patterns

We disaggregate the baseline specification by broad economic sector, estimating the Language Intensity × Post interaction separately for productivity growth in industry, services, and agriculture. Table B (Panel A) reports the results.

The sectoral pattern is consistent with our theoretical framework. Industry exhibits the strongest and most precisely estimated language effect. This is plausible: manufacturing and industrial production rely heavily on technical documentation, code-based automation, global value chain communication, and engineering problem-solving—all domains where English-language AI tools offer immediate productivity complementarities. Workers in these

sectors interact frequently with technical content that is overwhelmingly produced and indexed in English, and generative AI can directly accelerate tasks such as code generation, quality control documentation, and supply chain optimization.

Agriculture shows no discernible effect, consistent with the sector's limited digitization and the predominantly manual nature of agricultural work in most developing economies. The marginal productivity of AI tools is low when the binding constraints on output are land, weather, and physical labor rather than information processing.

The services estimate is positive but imprecisely estimated. This likely reflects well-known measurement difficulties in service-sector productivity rather than an absence of underlying effects. Service output is notoriously hard to measure at the cross-country level, and the heterogeneity within services—from financial consulting (highly AI-amenable) to personal care (minimally so)—introduces noise that aggregate data cannot resolve. Future research with more granular sectoral data could usefully decompose the services category to identify which subsectors drive AI productivity gains.

B.2 Gender and Labor Market Restructuring

We further examine whether AI-driven productivity gains are accompanied by shifts in labor market structure, focusing on gender differences. For high-Language Intensity economies, we estimate the effect of the ChatGPT shock on female and male labor market outcomes separately. Table B (Panel B) reports the results.

A suggestive pattern emerges in high-Language Intensity economies: female wage employment declines modestly while female self-employment rises by a comparable magnitude, with no significant change in overall female labor force participation. Male outcomes show no corresponding shifts. This pattern is consistent with recent evidence that ChatGPT adoption exhibits significant gender heterogeneity in labor market outcomes (Humlum and Vestergaard, 2025). One interpretation is that AI enables women to shift from traditional wage employment into platform-based work, freelancing, or AI-assisted entrepreneurship—

Table B1: Heterogeneous Effects by Sector and Gender

Panel A: Sectoral Heterogeneity

	(1) Agriculture	(2) Industry	(3) Services
Language Intensity × Post	0.032	0.187**	0.056
	(0.069)	(0.084)	(0.043)
Observations	1,144	1,144	1,144
R^2	0.003	0.079	0.057

Panel B: Gender Mechanisms

	(1) LFPR	(2) Wage Emp	(3) Self-Emp	(4) Emp Ratio
Language Intensity × Post	−0.135	−0.226*	0.226*	−0.075
	(0.157)	(0.120)	(0.120)	(0.189)
Observations	1,509	1,344	1,344	1,509
R^2	0.983	0.998	0.998	0.982
Country FE: Yes for all specifications		Year FE: Yes for all specifications		

Notes: Standard errors clustered at country level in parentheses. Panel A shows sectoral productivity growth (pp). Panel B shows female labor market outcomes. Industry shows strongest language effect (0.187**). Gender results suggest shift from wage to self-employment. Country and year FE in all regressions. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

sectors where language barriers historically limited participation and where generative AI tools reduce entry costs. In linguistically compatible economies, women may be among the first to leverage AI for occupational transitions that were previously inaccessible.

We emphasize several caveats. First, the magnitudes are modest and the estimates are imprecise, reflecting both the short post-treatment window and the inherent difficulty of measuring labor market composition shifts at the cross-country level. Second, we cannot distinguish between voluntary occupational transitions (women choosing self-employment enabled by AI) and involuntary displacement (AI automating female-dominated wage jobs). Third, the absence of effects for men does not imply men are unaffected by AI; it may instead reflect that male-dominated sectors (e.g., construction, heavy industry) are less immediately amenable to AI-driven restructuring, or that male labor market adjustments operate on different margins not captured by our outcome variables. These patterns point toward productive directions for future work with individual-level panel data that can trace worker-level transitions across occupations and employment types.

C Appendix: Model Derivations

This appendix provides the full derivation of the equilibrium and balanced growth path for the model presented in Section 2. The derivation closely follows Acemoglu et al. (2026); the key novelty is the decomposition of the adoption rate into conventional and AI-augmented channels.

C.1 Production and Pricing

Each country i produces a final good using a continuum of intermediate varieties:

$$Y_{it} = Q_{it}^{\frac{\sigma-2}{\sigma-1}} \left(\int_0^{Q_{it}} y_{jt}^{\frac{\sigma-1}{\sigma}} \, dj \right)^{\frac{\sigma}{\sigma-1}}, \tag{C.1}$$

where Q_{it} is the measure of available intermediates and $\sigma > 1$ is the elasticity of substitution. Each intermediate j is produced by a monopolist using one unit of labor per unit of output, so marginal cost equals the wage w_{it}.

The final good producer's demand for intermediate j is obtained from the first-order condition:

$$y_{jt} = Q_{it}^{\frac{\sigma-2}{\sigma-1}} \cdot p_{jt}^{-\sigma} \cdot P_{it}^{\sigma-1}, \tag{C.2}$$

where p_{jt} is the price of intermediate j and P_{it} is the ideal price index. Each monopolist maximizes $(p_{jt} - w_{it})y_{jt}$, yielding the standard constant-markup pricing rule:

$$p_{jt} = \frac{\sigma}{\sigma - 1} \cdot w_{it}. \tag{C.3}$$

In symmetric equilibrium, all intermediates are priced identically and produced in equal quantities. Substituting symmetric quantities into (C.1) and normalizing the wage to 1 (labor is the numéraire), we obtain:

$$Y_{it} = Q_{it}. \tag{C.4}$$

Output per worker equals the number of intermediate varieties. Per-intermediate monopoly profit is:

$$\pi = \frac{1}{\sigma}. \tag{C.5}$$

C.2 Technology Adoption

The world technology frontier Q_t^W grows at exogenous rate $g_W > 0$. Country i's relative technology position is $q_{it} \equiv Q_{it}/Q_t^W \in (0, 1]$.

A firm controlling n intermediates can invest $\Gamma \cdot Z$ units of the final good to adopt new varieties at flow rate $B_i \cdot Z$, where:

$$B_i = \theta_i(1 + \lambda_i \mu) q_i^{-\varepsilon}. \tag{C.6}$$

Here $\theta_i > 0$ is absorptive capacity, $\lambda_i \in [0,1]$ is linguistic compatibility with AI training data, $\mu \geq 0$ is the AI productivity parameter, and $\varepsilon > 0$ is the distance-to-frontier elasticity. When $\mu = 0$, this reduces to the standard AAJ adoption rate $B_i = \theta_i q_i^{-\varepsilon}$.

Define effective absorptive capacity as $\tilde{\theta}_i \equiv \theta_i(1 + \lambda_i \mu)$, so that $B_i = \tilde{\theta}_i q_i^{-\varepsilon}$.

C.3 Firm Value Function and Optimal Adoption

A firm controlling n intermediates earns flow profit $n \cdot \beta_i \pi$ (where β_i is the fraction surviving institutional expropriation) and chooses adoption effort Z to maximize its value. The firm's flow payoff net of adoption cost is:

$$n \cdot \beta_i \pi - \Gamma \cdot Z + B_i \cdot Z \cdot v_i \cdot n, \tag{C.7}$$

where v_i is the value per intermediate and $B_i \cdot Z \cdot n$ is the flow of new intermediates. The last term reflects that each new intermediate has value v_i and the total adoption rate scales with n (each existing variety generates opportunities to adopt related ones).

Defining $z_i \equiv Z/n$ (adoption effort per intermediate), the Hamilton-Jacobi-Bellman equation for the per-intermediate value is:

$$r_i v_i - \dot{v}_i = \beta_i \pi - \Gamma z_i + \tilde{\theta}_i q_i^{-\varepsilon} z_i v_i, \tag{C.8}$$

where r_i is the interest rate in country i.

The first-order condition for optimal z_i is:

$$\frac{\partial}{\partial z_i}\left[-\Gamma z_i + \tilde{\theta}_i q_i^{-\varepsilon} z_i v_i\right] = 0 \quad \Longrightarrow \quad \tilde{\theta}_i q_i^{-\varepsilon} v_i = \Gamma. \tag{C.9}$$

This pins down the per-intermediate value:

$$v_i = \frac{\Gamma}{\tilde{\theta}_i q_i^{-\varepsilon}}. \tag{C.10}$$

Intuitively, the value of an intermediate is higher when adoption is harder (low $\tilde{\theta}_i$, high q_i), because the scarcity rent is larger.

C.4 Law of Motion for Relative Technology

The growth rate of intermediates is $\dot{Q}_i/Q_i = B_i z_i(1-c_i)$, where $c_i \equiv C_i/Q_i$ is normalized consumption and $(1-c_i)$ captures the fraction of output available for adoption investment after consumption. Using $q_i \equiv Q_i/Q^W$:

$$\frac{\dot{q}_i}{q_i} = \frac{\dot{Q}_i}{Q_i} - g_W = \tilde{\theta}_i q_i^{-\varepsilon} z_i(1-c_i) - g_W. \tag{C.11}$$

Substituting the optimal z_i from the FOC (which implies $z_i = \beta_i \pi/[\Gamma(\rho + g_i)]$ on the BGP, but more generally is determined by the HJB), and using the free-entry condition to eliminate z_i, we obtain:

$$\frac{\dot{q}_i}{q_i} = \frac{\tilde{\theta}_i q_i^{-\varepsilon}(1-c_i)}{\Gamma} - g_W. \tag{C.12}$$

C.5 Euler Equation

The representative household maximizes $\int_0^\infty e^{-\rho t} \ln c_i \, dt$. The Euler equation in terms of normalized consumption c_i is:

$$\frac{\dot{c}_i}{c_i} = r_i - \rho - g_i, \tag{C.13}$$

where $g_i = \dot{Q}_i/Q_i$ is the domestic growth rate. Using the asset pricing relation (from the HJB at steady state) and substituting, this becomes:

$$\frac{\dot{c}_i}{c_i} = \frac{\tilde{\theta}_i q_i^{-\varepsilon}}{\Gamma}\left[\beta_i \pi + (\varepsilon - 1)(1-c_i)\right] - \rho - \varepsilon g_W. \tag{C.14}$$

Equations (C.12) and (C.14) form a two-dimensional dynamical system in (q_i, c_i).

C.6 Balanced Growth Path

On the balanced growth path, $\dot{q}_i = \dot{c}_i = 0$ and each country grows at the frontier rate g_W. Setting (C.12) to zero:

$$\frac{\tilde{\theta}_i (q_i^*)^{-\varepsilon}(1 - c_i^*)}{\Gamma} = g_W. \tag{C.15}$$

Setting (C.14) to zero:

$$\frac{\tilde{\theta}_i (q_i^*)^{-\varepsilon}}{\Gamma} [\beta_i \pi + (\varepsilon - 1)(1 - c_i^*)] = \rho + \varepsilon g_W. \tag{C.16}$$

From (C.15), we have:

$$\frac{\tilde{\theta}_i (q_i^*)^{-\varepsilon}}{\Gamma} = \frac{g_W}{1 - c_i^*}. \tag{C.17}$$

Substituting into (C.16):

$$\frac{g_W}{1 - c_i^*} [\beta_i \pi + (\varepsilon - 1)(1 - c_i^*)] = \rho + \varepsilon g_W. \tag{C.18}$$

Expanding and solving for c_i^*:

$$g_W \beta_i \pi + (\varepsilon - 1) g_W (1 - c_i^*) \cdot \frac{1}{1 - c_i^*} \cdot (1 - c_i^*) = (\rho + \varepsilon g_W)(1 - c_i^*). \tag{C.19}$$

Simplifying:

$$g_W \beta_i \pi + (\varepsilon - 1) g_W = (\rho + \varepsilon g_W)(1 - c_i^*) - (\varepsilon - 1) g_W (1 - c_i^*) + (\varepsilon - 1) g_W (1 - c_i^*). \tag{C.20}$$

More directly, from the asset pricing condition on the BGP. The per-intermediate value satisfies both (C.10) and the BGP asset pricing condition:

$$v_i = \frac{\beta_i \pi}{\rho + g_W}, \tag{C.21}$$

where $\rho + g_W$ is the effective discount rate on the BGP (the interest rate equals $\rho + g_W$ when

consumption grows at rate g_W).

Setting (C.10) equal to (C.21):

$$\frac{\Gamma}{\tilde{\theta}_i (q_i^*)^{-\varepsilon}} = \frac{\beta_i \pi}{\rho + g_W}. \tag{C.22}$$

Solving for q_i^*:

$$\boxed{q_i^* = \left(\frac{\tilde{\theta}_i \beta_i \pi}{\Gamma(\rho + g_W)} \right)^{1/\varepsilon} = \left(\frac{\theta_i (1 + \lambda_i \mu) \cdot \beta_i \pi}{\Gamma(\rho + g_W)} \right)^{1/\varepsilon}}. \tag{C.23}$$

This is equation (4) in the main text.

C.7 Proof of Saddle-Path Stability

The dynamical system (C.12)–(C.14) has a unique steady state (q_i^*, c_i^*). Linearizing around this steady state, the Jacobian has one positive and one negative eigenvalue (the same structure as in AAJ Proposition 2), ensuring saddle-path stability. The proof is identical to that in Acemoglu et al. (2026) with θ_i replaced by $\tilde{\theta}_i = \theta_i(1 + \lambda_i \mu)$, since the functional form of the dynamical system is unchanged. The negative eigenvalue governs the speed of convergence along the saddle path.

C.8 Proofs of Propositions

Proof of Proposition 1. From (C.23), the ratio of post-AI to pre-AI steady states is:

$$\frac{q_i^{*,\text{post}}}{q_i^{*,\text{pre}}} = \left(\frac{\theta_i (1 + \lambda_i \mu) \beta_i \pi / [\Gamma(\rho + g_W)]}{\theta_i \beta_i \pi / [\Gamma(\rho + g_W)]} \right)^{1/\varepsilon} = (1 + \lambda_i \mu)^{1/\varepsilon}. \tag{C.24}$$

All terms except λ_i and μ cancel. □

Proof of Proposition 2. Setting $\lambda_i = 0$ in (C.23):

$$q_i^{*,\text{post}} = \left(\frac{\theta_i(1 + 0 \cdot \mu)\beta_i\pi}{\Gamma(\rho + g_W)}\right)^{1/\varepsilon} = \left(\frac{\theta_i\beta_i\pi}{\Gamma(\rho + g_W)}\right)^{1/\varepsilon} = q_i^{*,\text{pre}}. \tag{C.25}$$

The steady state is unchanged regardless of θ_i, β_i, or any other parameter. □

Proof of Proposition 3. Consider countries R and P with $\theta_R\beta_R > \theta_P\beta_P$ (so $q_R^{*,\text{pre}} > q_P^{*,\text{pre}}$) and $\lambda_R < \lambda_P$. In the post-AI steady state, $q_P^{*,\text{post}} > q_R^{*,\text{post}}$ if and only if:

$$\theta_P\beta_P(1 + \lambda_P\mu) > \theta_R\beta_R(1 + \lambda_R\mu). \tag{C.26}$$

Expanding:

$$\theta_P\beta_P + \theta_P\beta_P\lambda_P\mu > \theta_R\beta_R + \theta_R\beta_R\lambda_R\mu. \tag{C.27}$$

Rearranging:

$$\mu(\theta_P\beta_P\lambda_P - \theta_R\beta_R\lambda_R) > \theta_R\beta_R - \theta_P\beta_P. \tag{C.28}$$

The right-hand side is positive by assumption (R is initially richer). A finite positive threshold $\bar{\mu}$ exists if and only if the left-hand side coefficient is positive:

$$\theta_P\beta_P\lambda_P > \theta_R\beta_R\lambda_R. \tag{C.29}$$

When this holds:

$$\bar{\mu} = \frac{\theta_R\beta_R - \theta_P\beta_P}{\theta_P\beta_P\lambda_P - \theta_R\beta_R\lambda_R}. \tag{C.30}$$

For all $\mu > \bar{\mu}$, the poor country overtakes the rich country. The condition $\theta_P\beta_P\lambda_P > \theta_R\beta_R\lambda_R$ requires that the poor country's linguistic advantage be sufficiently large relative to the rich country's traditional advantage—precisely the case for pairs such as Nigeria ($\lambda = 1.00$, $\theta\beta$ low) versus Japan ($\lambda = 0.11$, $\theta\beta$ high). □

C.9 Variance Decomposition

Taking logs of (C.23):

$$\ln q_i^{*,\text{post}} = \frac{1}{\varepsilon} \ln \left(\frac{\theta_i \beta_i \pi}{\Gamma(\rho + g_W)} \right) + \frac{1}{\varepsilon} \ln(1 + \lambda_i \mu). \tag{C.31}$$

The first term equals $\ln q_i^{*,\text{pre}}$. Therefore:

$$\ln q_i^{*,\text{post}} = \ln q_i^{*,\text{pre}} + \frac{1}{\varepsilon} \ln(1 + \lambda_i \mu). \tag{C.32}$$

Taking the cross-country variance:

$$\text{Var}(\ln q_i^{*,\text{post}}) = \text{Var}(\ln q_i^{*,\text{pre}}) + \frac{1}{\varepsilon^2} \text{Var}(\ln(1 + \lambda_i \mu)) + \frac{2}{\varepsilon} \text{Cov}(\ln q_i^{*,\text{pre}}, \ln(1 + \lambda_i \mu)). \tag{C.33}$$

This is equation (8) in the main text. The second term is unambiguously positive. The covariance term is empirically small because $\text{Corr}(\lambda_i, \ln q_i^{*,\text{pre}}) \approx -0.09$, implying that AI approximately unconditionally increases cross-country inequality. □

D Appendix: Task Texts

This appendix reproduces one English-language instance of each business decision task (numerical values vary across instances; non-English versions are produced by translating the template while holding all numerals fixed).

Production planning (linear programming)

```
PRODUCTION PLANNING MEMO. This month our small factory makes two products,
the Standard and the Artisan. After the recent change in packaging costs,
each Standard nets $50 in profit and each Artisan nets $20. Both pass through
```

our CNC cell and then final assembly. On the CNC, one Standard takes 2 machine-hours and one Artisan takes 3; in assembly, a Standard needs 3 hours and a Artisan needs 2. Because of scheduled maintenance this month we have only 300 CNC-hours and 400 assembly-hours available. Sales advises that the boutique channel cannot absorb more than 100 units of the Standard. (For the record, the $4000 we spent last year on tooling is a sunk cost and should not influence this decision.) How many of each product should we produce this month to maximize total profit? End your reply with exactly: ANSWER: <units of Standard>, <units of Artisan>

Capacity expansion / shadow price

OPERATIONS MEMO. Artisan and Premium are this month's two products. After costs, each Artisan contributes $50 in profit and each Premium $40. Both run through the CNC cell and then final assembly: a Artisan takes 2 CNC-hours and 3 assembly-hours, while a Premium takes 2 CNC-hours and 2 assembly-hours. This month we have 300 CNC-hours and 500 assembly-hours available, and sales can place at most 120 units of the Artisan. A contractor has now offered to rent us 60 additional CNC-hours this month for a flat fee of $720. By how much would accepting that offer change our maximum total profit for the month, net of the fee? End your reply with exactly: ANSWER: <net change in dollars, negative if worse off>

Value of information (decision tree + Bayes)

CAPITAL PROJECT MEMO -- for the investment committee. Over the past few weeks we have been weighing how, or whether, to bring Kestrel Labs's new product line to market, and the committee has asked for a single recommendation. Three paths are on the table: commit now to a full large-scale plant, start more cautiously with a smaller line, or simply hold off for the time being

(which of course neither costs nor earns us anything). What makes the call hard is how much hinges on demand, which none of us can pin down yet. If the line really catches on, finance reckons the large plant would ultimately clear about $700, whereas the smaller, capacity-constrained line would only bring in around $320. If demand instead comes in soft, the picture flips: the large plant turns into a burden and would land at roughly $-400, while the smaller line would come in near $180 (read any negative figure as a loss). Marketing currently puts the probability of that strong-demand scenario at about 40%. There is also one more option worth raising: before committing to anything, we could pay $70 to have the same market-research firm we used last year run a demand study first. It is far from a crystal ball, though. Looking back over their track record, among the launches that genuinely went on to succeed their report had read 'favourable' about 80% of the time; but even among the ones that ultimately flopped, it had still come back 'favourable' around 10% of the time. Weighing all of this together, which single course of action maximises our expected value?

A: Skip the study and commit to the large plant now.

B: Skip the study and go with the smaller line now.

C: Pay for the study first, then build large if it reads 'favourable' and the smaller line if 'unfavourable'.

D: Pay for the study first, then build large if it reads 'favourable' and hold off entirely if 'unfavourable'. End your reply with exactly: ANSWER: <A, B, C, or D>

Market-entry strategy

STRATEGY MEMO -- Market Entry Decision. Verda Foods will enter exactly one new market this year. The board has been explicit about our strategy: we enter only markets that are a strong fit with our core competencies, and

we will not enter heavily regulated sectors -- no matter how fast they are growing. Four markets cleared the initial screen. Our analysts summarized each:

Market A: only a partial fit with our core competencies; a moderately regulated sector; modest growth prospects; contested competition.

Market B: a strong fit with our core competencies; a moderately regulated sector; solid growth prospects; largely open competition.

Market C: a strong fit with our core competencies; a heavily regulated sector; explosive growth prospects; largely open competition.

Market D: a weak fit with our core competencies; a lightly regulated sector; explosive growth prospects; largely open competition.

Given our stated strategic priorities, which single market should we enter?

End your reply with exactly: ANSWER: <A, B, C, or D>

Business communication (subtext)

INTERNAL NOTE. You asked your key supplier to confirm they can hit the product-launch deadline. Their reply: "Thanks so much for your patience here. I'll be candid -- the timeline you've laid out is certainly ambitious, and while we'd genuinely love to make it work, a couple of things on our side have turned out trickier than we'd hoped. We're fully committed to the partnership and will absolutely do our very best. Why don't we touch base again next month once things have firmed up?" For planning purposes, how should you treat the deadline?

A: They confirmed commitment and will do their best -- plan around on-time delivery.

B: This is a soft warning the deadline will slip -- line up a contingency supplier now.

C: They are asking to renegotiate price -- prepare a counteroffer.

D: They have declined the project entirely -- cancel the launch. End your

```
reply with exactly: ANSWER: <A, B, C, or D>
```

E Appendix: Robustness of the Calibration to r and ε

The bridge equation (12) implies that the regression slope of growth accelerations on standardized Language Intensity is governed by $r\mu/\varepsilon$. Cross-country DID pins down this composite to match $\hat{\beta} = 0.94$ pp; any combination (r, ε, μ) satisfying $r\mu/\varepsilon \approx 0.033$ produces the same fit. Quantitative magnitudes therefore scale with r, while ratios between countries or scenarios depend only on the composite and are invariant.

Table B2 reports the calibration across $r \in [0.04, 0.15]$, spanning the conditional convergence literature. The calibrated μ ranges from 0.91 to 0.24 and the high-λ steady-state gain from 80 to 22 percent, while the multilingual variance reduction stays in 53–56 percent and the UK/China gain ratio stays at 7.4–7.5. Table B3 reports the same exercise across $\varepsilon \in [0.8, 1.5]$ at $r = 0.10$: the calibrated μ adjusts proportionally, leaving counterfactual quantities essentially unchanged (UK gain 32.08% to 32.30%; dollar loss \$3.53T to \$3.79T). The paper's qualitative conclusions hold throughout both ranges; the baseline $r = 0.10$ is conservative within the r range.

F Appendix: Leapfrogging Within the Colonial Sample

This appendix reports the within-colonial-sample leapfrogging estimates summarized in Section 7.2. Restricting attention to former colonies holds colonial-institutional legacies approximately constant; column 1 shows poor British colonies gaining 2.55 percent relative to poor non-British ones, column 2 a negative triple interaction with income (the largest gains in the poorest colonies), and column 3 the four-quadrant ranking replicated within the colonial sample.

Table B2: Robustness of Calibration to Convergence Rate r

r	0.04	0.05	0.10†	0.13	0.15
Source	MRW	Islam	CEL	Mid-high	Upper
Calibrated parameter:					
μ	0.914	0.727	**0.359**	0.276	0.239
Country-level steady-state gain (%):					
UK / Nigeria / Ghana ($\lambda = 1$)	80.4	64.4	**32.2**	24.8	21.5
China ($\lambda = 0.13$)	10.9	8.7	**4.3**	3.3	2.9
Japan ($\lambda = 0.11$)	8.7	7.0	**3.4**	2.6	2.3
Korea ($\lambda = 0.02$)	1.3	1.1	**0.5**	0.4	0.4
Aggregate inequality and welfare:					
Annual output loss ($T)	9.21	7.34	**3.64**	2.79	2.42
Δ Gini (pop-wt.)	0.050	0.041	**0.021**	0.017	0.014
Invariant (qualitative) results:					
Multilingual var. reduction (%)	56	56	**54**	53	53
UK / China gain ratio	7.4	7.4	**7.5**	7.5	7.5

Notes: Each column corresponds to a value of r from the cross-country growth literature: MRW = Mankiw et al. (1992); Islam = Islam (1995); CEL = Caselli et al. (1996). For each r, μ is recalibrated to match $\hat{\beta} = 0.94$ pp. † Baseline.

Table B3: Robustness to Distance-to-Frontier Elasticity ε

ε	0.8	0.9	1.0	1.1†	1.2	1.3	1.5
μ	0.251	0.286	0.322	**0.359**	0.398	0.437	0.518
UK / Nigeria gain (%)	32.30	32.26	32.23	**32.20**	32.17	32.14	32.08
China gain (%)	4.14	4.19	4.24	**4.29**	4.34	4.39	4.49
Annual output loss ($T)	3.53	3.57	3.60	**3.64**	3.68	3.71	3.79
Δ Gini	0.0214	0.0214	0.0214	**0.0214**	0.0215	0.0215	0.0216

Notes: Sensitivity of headline counterfactual quantities to ε, holding $r = 0.10$. † Baseline value, following Acemoglu et al. (2026).

Table B4: Linguistic Leapfrogging: Poor English-Speaking Colonies

	Productivity Growth		
	(1)	(2)	(3)
	Poor British	Lang × Income	4-Way Matrix
Rich British Colony × Post			−0.827
			(0.848)
Poor British Colony × Post			2.353*
			(1.297)
			(Unexpected Winner)
Rich Non-British Colony × Post			−0.023
			(0.984)
			(Unexpected Loser)
Poor British Colony × Post	2.553**		
	(1.160)		
Language Intensity × Post		6.510***	
		(1.913)	
Lang × Post × GDP (2019)		−0.451***	
		(0.163)	
Log GDP per capita	17.191***	19.074***	17.170***
	(4.640)	(2.626)	(4.692)
Internet penetration (%)	1.087	−1.207	1.025
	(3.745)	(5.486)	(3.758)
Observations	673	268	673
Adjusted R^2	0.390	0.649	0.389
Country FE	Yes	Yes	Yes
Year FE	Yes	Yes	Yes

Notes: Dependent variable is annual labor productivity growth rate (percentage points). Poor/rich classification based on 2019 GDP per capita (pre-treatment) using median split within the colonial sample.

All specifications include country and year fixed effects. Standard errors in parentheses, clustered at country level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.