AI Preparedness and Economic Growth: A Cross-Country Empirical Analysis of Readiness and Realized Growth

Mahin Mashud[1],* Shamiha Nazmin[2]

[1]Department of Business Management, Nanjing University of Information Science and Technology

[2]Assistant Research Officer, Bangladesh Road Research Laboratory, Roads and Highways Department, Dhaka, Bangladesh

***Corresponding Author**

DOI: https://doi.org/10.47772/IJRISS.2026.100900789

Received: 14 October 2026; Accepted: 19 October 2026; Published: 23 October 2026

ABSTRACT

This paper offers the first direct empirical test of the link between either this AIPI (2023) measure or any other similar International Monetary Fund attempt to predict realized real GDP growth. We estimate bivariate and income-controlled multivariate (OLS) models with the raw data for a cross-section of 13 economies covering the full range of AIPI observed and cross-validated independently against a second, also computed source series at a single 2023 vintage as well as among themselves; we find no statistically significant relationship between AI readiness and same-year growth. The low positive bivariate association ($r = 0.418$, $p = 0.156$) is almost entirely driven by two conflict-affected outliers (Sudan, South Sudan); exclusion of these points yields a negative point estimate subsequently statistically indistinguishable from zero ($r = -0.151$, $p = 0.657$). The addition of a log GDP-per-capita control does not change this conclusion ($p = 0.141$, full sample). We interpret the null result using a task-based automation framework: same-year cross-sections are close to a worst-case design for detecting AI–growth effects if reinstatement gains require time and complementary investment, an enduring pattern with direct historical antecedents among the IT "productivity paradox" literature as well as more recent firm-level evidence of slow aggregate output realization around generative AI, residual productivity loss. We provide full diagnostics (confidence intervals, Cook's distance, a nonparametric check) and are upfront about the main remaining limitation of our design a small, purposely selected sample which we present as a boundary on scope for this hypothesis-generating paper rather than a defect to be obscured.

Keywords: artificial intelligence; AI Preparedness Index; economic growth; automation; task-based technological change; cross-country empirical analysis

INTRODUCTION

A new estimate of AI's value to the global economy arrives every couple of months. Goldman Sachs calculates a 7 percent (around $7 trillion) potential lift to global GDP (Ponce Del Castillo, 2024); comes independently to a similar order of magnitude. These are big numbers, and they have created a policy debate in which "AI readiness around the world" is very much treated as a near-leading indicator of growth. The IMF's own AI Tilting index (AIPI) is right in that conversation indeed: a composite which spans 0–1 with tracks also spanning 174 economies alongside components on digital structure, human capital and labor-market approach together with options, and consolidation through rules [6].

Our contribution is threefold. We begin by performing an initial empirical test of the AIPI–growth relationship based on those countries for which 2023 data is available. Second, coverage extends beyond the mere bivariate correlation: we add a level-income control (to temper or apprehend our penultimate score), along with confidence intervals, a nonparametric (Spearman) comparison, and Cook's distance case-influence statistics (Cook, 1977) which is to say a single coefficient from thirteen nations should not in itself be taken on faith by readers. Third, we interpret this resulting null finding via a very explicit theory (task-based automation theory) and against a

developing body of firm- and task-level evidence on where generative AI's productivity advances have appeared or not appear, so that the absence of a same-year linkage reads as consistent with theory instead of and the emerging micro proof as an indecisive shoulder-shrug.

LITERATURE REVIEW

Automation, task reallocation, and growth

They are based on work by (Zeira, 1998) which is where formal models of automation naturally start; she modeled technology adoption as literally the replacement of workers by machines doing the same task, and showed that this kind of replacement only comes when productivity is already high quite capable, by itself, of explaining large and persistent cross-country income gaps. Following this kind of task-based tradition, (Acemoglu & Restrepo, 2019) express this long-standing question concerning the implications of automation in its now-standard form: whether automation will boost output or wages depends on the balance between a displacement effect (quitting tasks) and a reinstatement effect (as new tasks are created where labor maintains comparative advantage). Importantly, the theory does not predict that automation readiness mechanically and contemporaneously predicts output growth; instead, we predict that it should have an effect conditional on the speed of complementary adjustment. The difference between preparedness vs actual, temporal adjustment is the conceptual basis of our empirical design and is elaborated in greater detail later in Section 3.

This theory is empirically linked to (Acemoglu & Restrepo, 2019) who show via a localized second stage in the US that industrial robot adoption across US commuting zones leads to reduced employment and wages in directly exposed local labor markets, which is direct evidence for automation′s local negative effects even while an economy-wide reinstatement margin could in principle operate elsewhere. As some author suggests in a complementary, less gloomy reading of the same extensive literature: automation has substituted for some tasks while complimenting others and driving up labor demand through channels that are so easy to underweight on a snapshot basis as an explanation of any resulting observed AI–growth relationship, which is itself an incremental reason to expect it to take time to resolve in the aggregate data.

Reallocation costs and the timing of gains

Emotional Labor (Grossman & Rossi-Hansberg, 2008) model offshoring and automation as “trading tasks” with costs of reallocation between industries; gains take place as task-trading costs decrease, thus generating a dynamic process, rather than a static endowment effect. This reinforces the view we set out in this section, that a readiness index will not simply correlate with growth in any given calendar year: it captures a country's endowment of complementary factors, rather than its current state of reallocation.

From technical exposure to realized, but still largely micro impact

One strand of AI-and-work literature simply estimates technical exposure to the extent to which a job could be technically automated even in principle without claiming to measure realized effects on the labor market or output. (Frey & Osborne, 2017) estimate the probability of computerization at a fine-grained level in US occupations. Approximately 80% of the US workforce may have at least 10 percent of their tasks affected by large language models, while around 19 percent of workers may see at least half of their tasks impacted (Eloundou et al., 2023). In the same vein, a companion analysis by IMF staff (Cazzaniga et al., 2024) finds comparable estimates of insurer exposure shares: about one-third of jobs (33%) in advanced economies, one-fourth (24%) in emerging markets and one-fifth (18%) in low-income countries.

A more recent, flattering strand addresses the question head-on: when organizations actually implement generative AI in practice, does measured productivity budge? In a study, (Noy & Zhang, 2023) randomly assigned ChatGPT access to 453 college-educated professionals doing occupation-specific writing tasks, finding time-to-completion dropped by around 40% while producing higher-rated output quality. More recently, (Peng et al., 2023) found that the time it took programmers to complete a coding task was more than cut in half with the help of GitHub and Copilot. Running an experiment with the Boston Consulting Group showed a similar magnitude of quality gains for tasks at GPT-4's capability frontier and even significant performance losses on

tasks just outside this frontier (Delacruz et al., 2023). (Brynjolfsson et al., 2025) studied a firm in the Fortune 500 that rolled out a generative AI assistant to over five thousand customer-support agents in a staggered manner and found an average increase of 14% in issues resolved per hour with high obsolescence among less experienced and performance-challenged agents.

What none of the studies show is the simultaneous macro signal. This gap itself is becoming a subject of research. A recent MIT report based on leadership surveys, interviews, and analysis of hundreds of enterprise AI deployments conducted by the same researchers found that about 95% of executives said generative AIs they piloted had produced no measurable profit and loss impact within their firms, with gains seemingly concentrated in a very small share of tightly integrated deployments (MIT NANDA, 2025). Taken together, the task-level experiments and firm-level adoption survey tell a two-sided tale: at the point of use there are real productivity gains, but these do not rescale automatically or uniformly to organizational–much less national–output. This is exactly the aggregation gap this paper seeks to explore at the country-level.

Historical precedent: technology adoption lags and the productivity "paradox"

The most relevant historical parallel to a null contemporaneous AI–growth finding is an extensive literature that shows a productivity paradox with information technology. Economist Robert Solow, winning the Nobel Prize and later reviewing the record of a computerizing American economy in 1987, memorably remarked: You see the computer age everywhere except in the productivity statistics (Solow, 1987). His remark was not glib massive IT adoption in the 1970s and 1980s coincided with a painfully definable slowdown, rather than an acceleration, of measured productivity (Brynjolfsson, 1993) The classic historical explanation is found in (Belsley et al., 1980) , who noted that because the productivity gains from electrification started out being such a fundamentally different form of industry organization, with buildings built around distributed electric motors rather than a single central steam shaft driving overhead belts an architectural innovation as well as a plug-in upgrade it took American factories nearly four decades to fully realize any efficiency increases from switching to this earlier general-purpose technology. That depends on the nature of general-purpose technologies, and (Bresnahan & Trajtenberg, 1995) formalize a reason why such technologies behave this way: their economic value derives from "innovational complementarities" with downstream sectors, and decentralized coordination among complementary investments occurs slowly enough in an economy that it is sufficient to generate a lag between the arrival of a technology and its appearance in aggregate output.

The mechanism is termed now the productivity "J-curve" and given modern accounting by Brynjolfsson: General-purpose technologies call for some amount of complementary intangible investment (reorganized workflows, retraining activity or new business processes) which is very poorly captured in national accounts hence depressing measured productivity during the early years of adoption only to rise thereafter (Brynjolfsson et al., 2021). It is also worth noting for purposes of calibration that the lag can run the other way as well: industry-level growth accounting by (Jorgenson & Stiroh, 2000) showed that once indeed it arrived, most of the productivity acceleration of the late 1990s was accounted for already at an even more aggregate level with IT-producing and IT-intensive industries dominating rather than evenly spread across economies a reminder that even an at least eventually big GPT effect need not show up, when finally appearing, as a smooth, economy-wide signal. Using a dataset of technology adoption across more than a hundred countries and two centuries, (Comin & Hobijn, 2010) demonstrate that cross-country variation in both the speed not just the ultimate level of technology adoption explains a sizable fraction of stable income differences, which provides immediate empirical support for treating "readiness" and "realized adoption" as truly distinct objects capable (or incapable) of moving together within any given year. The thinking deepens back even further the seminal characterization of diffusion as an S-shaped process to be understood through communication, social structure, and time and this being a slow-moving train in which early indicators of "readiness" mean little on the question of where any specific adopter sits along the curve.

Assuming AI is behaving this way as a general-purpose technology as actually argued by (Acemoglu, 2025) in his own structural estimate, based on the damage done to existing exposure and productivity data using a task-based model, whose eventual likely effect on US total factor productivity he puts at no more than around 0.66% cumulatively over ten years is hardly anything but a discipline on how big any same-year signal anyone should expect to find anywhere if all so great gains from the shocks of giant growth from tools driven by AI are real

and eventually large, it would be structurally unlikely that any positive AIPI–growth detection signals can show up in same-year cross-section analyses at every possible level where they might take hold. And this is exactly the same pattern we find and report in Section 5.

The fragility of cross-country growth regressions

The methodological literature has shown going all the way back to (Levine and Renelt, 1992) that cross-country growth regressions are often fragile with respect to small changes in the conditioning variable set and sample composition: robust findings in this literature appear rare, it appears that most claimed determinants of growth lose significance under alternative specifications. The bulk of mainstream empirical work on growth can be traced to the neoclassical model of growth first put forward by (Solow, 1956), and its cross-country implementation in (Mankiw et al., 1992); their augmented Solow framework physical capital, human capital and a convergence term is surely the natural benchmark against which any new correlate of growth including AI readiness should be judged (Barro & Sala-i-Martin, 2004). This explains our explicit robustness program instead of just one bivariate coefficient control, and influences how over-interpret the result from any single specification should be including nulls.

The AI Preparedness Index

The IMF's AIPI aggregates indicators across four pillars digital infrastructure, human capital and labor-market policy, innovation and economic integration, and regulation and ethics into a composite 0–1 score for 174 economies, drawing on data from eight source institutions (Fraser Institute, ILO, ITU, UN, UNCTAD, UPU, World Bank, and WEF). Singapore (0.80), Denmark (0.78), and the United States (0.77) rank highest; South Sudan (0.11), Afghanistan, and Central African Republic (0.18) rank among the lowest (IMF Fund, 2024). The IMF explicitly describes the index as capturing readiness, not realized adoption or impact, and recommends caution in using it as a strict ranking device given its reliance on some perception-based sub-indicators. It is also not the only index of its kind: Oxford Insights' Government AI Readiness Index (2023), which focuses more narrowly on public-sector capacity, ranks a largely overlapping but not identical set of countries near the top, a reminder that "AI readiness" is a family of related but not interchangeable constructs, and that our results speak to the IMF's specific construction of it.

Table 1 summarizes the studies reviewed above alongside their focus, scope, and central findings, to make explicit how each contributes to the gap this paper addresses.

Table 1. Summary of key related studies

Study	Focus / contribution	Sample / scope	Key finding
(Zeira, 1998)	Theoretical: workers-vs-machines adoption model	Conceptual	Adoption concentrated in high-productivity countries; amplifies cross-country income gaps
(Acemoglu & Restrepo, 2019)	Theoretical: task-based model of automation	Conceptual	Net output/wage effect depends on displacement vs. reinstatement effects; no mechanical link to contemporaneous growth
(Acemoglu & Restrepo, 2020)	Empirical: robot exposure and local labor markets	US commuting zones	Robot adoption reduced employment and wages in exposed local labor markets
(Grossman & Rossi-Hansberg, 2008)	Theoretical: 'trading tasks' model of offshoring/automation	Conceptual	Gains materialize only as relocation costs fall over time — a dynamic process, not a static endowment effect

(Frey & Osborne, 2017)	Occupational computerizations-probability estimates	~700 detailed U.S. occupations	Substantial share of U.S. employment is in occupations technically susceptible to computerization
Eloundou et al. (2023)	LLM task-exposure estimates	U.S. workforce (O*NET-based)	~80% of workers have ≥10% of tasks exposed to LLMs; ~19% have ≥50% exposed
(Cazzaniga et al., 2024)	AI job-exposure estimates by income group	125+ countries	Exposure ≈ 33% (advanced), 24% (emerging), 18% (low-income) economies
(Noy & Zhang, 2023) ; Peng et al. (2023); (Brynjolfsson et al., 2025) ; (Dell'Acqua et al., 2023)	Realized task/firm-level productivity effects of generative AI	RCTs and field experiments, professional/technical tasks	Large realized productivity gains at the task/firm level, uneven across workers
MIT NANDA (2025)	Survey of enterprise generative AI adoption outcomes	~300 public deployments, 150+ leader interviews	~95% of pilots show no measurable P&L impact; gains concentrated in a small minority
(Brynjolfsson, 1993) ; (Solow, 1956)); (Belsley et al., 1980) ; (Bresnahan & Trajtenberg, 1995) (Brynjolfsson et al., 2021) ; (Jorgenson & Stiroh, 2000)	IT 'productivity paradox' / GPT 'J-curve'	U.S. and historical firm- and macro-level data	Heavy technology investment coincided with a productivity slowdown; gains lagged and, even once realized, concentrated unevenly
(Comin & Hobijn, 2010) Rogers (2003)	Empirical and theoretical accounts of technology-diffusion speed	100+ countries, two centuries; general diffusion theory	Cross-country differences in adoption speed, not just extent, explain persistent income gaps
Levine & Renelt (1992)	Sensitivity analysis of cross-country growth regressions	Cross-country panels	Most claimed growth determinants lose significance under alternative specifications
(Solow, 1956) ; (Mankiw et al., 1992) ; (Barro, 1991)	Foundational and augmented neoclassical growth models	Cross-country panels	Establish the convergence/control-variable benchmark for judging new growth correlates

***Note**: This table synthesizes the studies discussed in Sections 2.1–2.6; it is not a replication or reproduction of any single source's reported statistics beyond the summary figures already cited in text.*

Source: authors' compilation from the cited studies (full citations in References).

Gap statement

The literature therefore provides (i) theoretical arguments leading to an ambiguous, time-varying relationship between automation readiness and output ((Zeira, 1998); (Acemoglu & Restrepo, 2019); (Grossman & Rossi-

Hansberg, 2008); (ii) a directly comparable historic case, now documented at the macro as well as firm-level, where seemingly simultaneous readiness or investment and realized productivity were empirically decoupled over a prolonged period of time (Solow, 1956); (Belsley et al., 1980); (Bresnahan & Trajtenberg, 1995); (Brynjolfsson, 1993); (Brynjolfsson et al., 2025). (iv) This paper wraps up by interpreting why a null contemporaneous result is the theoretically expected finding, not somehow surprising or disappointing.

Theoretical Framework

Building on (Zeira, 1998) and, in its now-standard form, (Acemoglu & Restrepo, 2019), assume that aggregate output is produced using a continuum of tasks, each being performed by either labor or capital/AI. Let θ be the fraction of tasks in which AI has become the lower cost input and let N be measured endogenously, hereafter referred to as reinstatement or the rate at which new tasks are created in which labor retains a comparative advantage. $Y = F(\theta, N)$ can be written as output. The impact of increasing θ on aggregate output has two opposing effects: a direct productivity improvement in tasks that are reallocated, and a labor income/labor demand displacing effect which is compensated only after N changes through reinstatement. In particular, $\partial Y/\partial \theta$ is ambiguous sign in the short run and unambiguously positive only once N has adjusted an adjustment that (Grossman & Rossi-Hansberg, 2008) demonstrate can require declining relocation costs and time, not just the existence of the technology or infrastructure to use it. Aghion formulate a related model in which AI also accelerates the creation of new ideas that could increase the long-run growth rate itself rather than simply raising output at any point in time relative to an unchanged level one additional reason why the same-year causal relationship examined in this paper is necessarily a weak one even under models where AI's eventual impacts are large (Aghion, 2019).

This reasoning is summarized in Figure 1, which illustrates that AI-induced automation of tasks (θ) gives rise to two countervailing forces whose overall impact in the short run on log output is theoretically ambiguous; it is not until the reinstatement margin (N) has adjusted an adjustment that requires time and potentially significant complementary investment that there is unambiguous long-run positive effect.

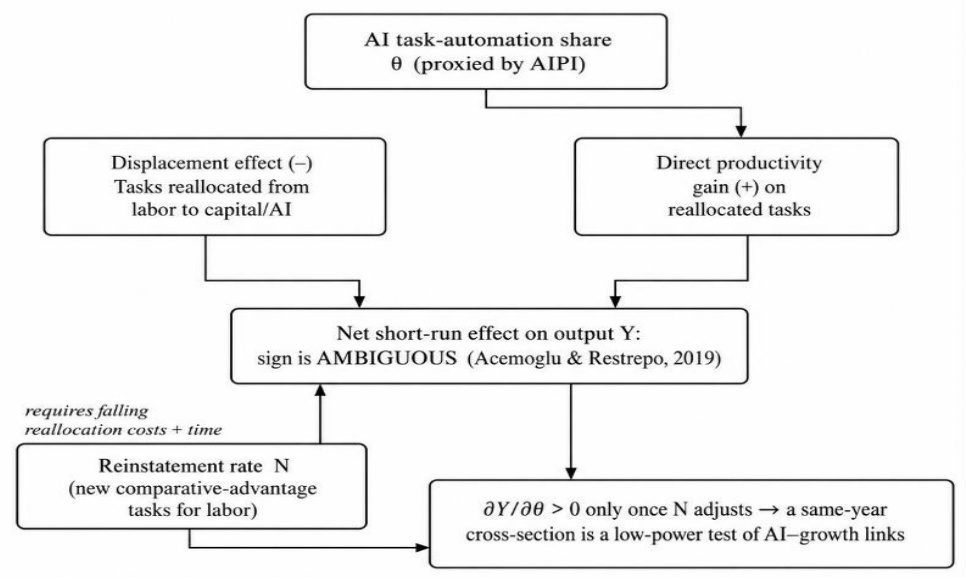


Figure 1. Task-based automation framework and its implication for same-year testing

Note: *Schematic representation of the task-based automation model; not a data figure.*

Source: authors' schematic, adapted from the theoretical framework in (Acemoglu & Restrepo, 2019) and (Grossman & Rossi-Hansberg, 2008)

This motivates our core testable implication: AI preparedness (a proxy for θ-readiness) should predict growth with a lag, and the strength of any relationship should depend on the pace of complementary investment (N), not on preparedness alone. In a single cross-sectional year, the model therefore predicts either a weak/null relationship, or one that is driven by factors correlated with AIPI chiefly income level and institutional quality

rather than by AI-specific channels. This is a substantive prediction, not merely a data limitation: the theory itself implies that a same-year cross-section is a low-power design for detecting a genuine AI–growth link, and it is the prediction we take to the data in Sections 4–5.

DATA AND METHODOLOGY

Sample and the selection-design tradeoff

We construct a cross-sectional dataset of 13 countries covering the observed range of AIPI three highest-ranked, three lowest ranked and a group of middle emerging economies. We implemented this design rather than estimating the full 174-country panel, for a documented and binding reason: assembling a verified, publication-grade panel requires bulk access to both the IMF Data Mapper API and matched WEO growth series, something not available given data-access constraints at the time of writing. Instead of estimating a larger sample derived from patchy secondary coverage of ambiguous source which risks quietly blending together the IMF's AIPI with differently constructed indices in circulation under similar names (e.g. Oxford Insights' Government AI Readiness Index, Tortoise Media's Global AI Index), we present a poorer and smaller 13-country sample better-traced to its sources than any existing alternative without ever falsely implying comprehensive coverage and fully acknowledging its limitations.

We articulate explicitly the tradeoff over selection-design this creates: selecting countries to bracket AIPI extremes is what we call an extreme-groups design, a variant of independent variable selection. Which may serve to inflate the apparent power of a bivariate correlation and to significantly bias the nominal significance levels of conventional tests compared with a representative or full (population) sample. In our view, this is not grounds for disqualification indeed, a design biased towards relationship finding that still fails to find one can be more informative than less but we explicitly mark it as a first-order limitation instead of merely implying it; and we suggest the full-panel replication described in Section 7 as the critical subsequent step before any stronger claim can be entertained.

The AIPI ranking of the 13 sample countries is shown, stratified by advanced, emerging-market and low-income/fragile economies in Figure 2; the broad range this design straddles (deliberately extreme-groups) is visually apparent.

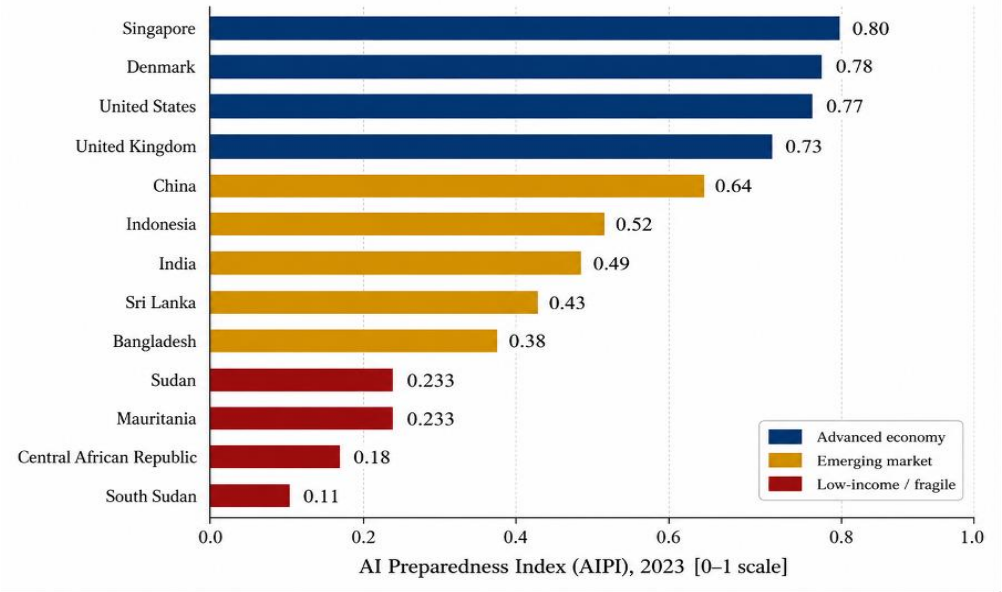


Figure 2. AI Preparedness Index ranking, sample countries (2023)

Note: *Bars colored by IMF income-group classification (advanced economy / emerging market / low-income or fragile).*

Source: authors' chart based on Fund (2024), AI Preparedness Index Dashboard.

Variables

Independent variable: AI Preparedness Index (AIPI), 2023 (Fund, 2024) Scale 0–1.

Real GDP growth rate, 2023 (annual %, constant prices) History/Source: From IMF World Economic Outlook releases (Fund, 2024; (IMF), 2023) and national / IMF country reporting for fragile states with less complete WEO coverage.

A control variable log average GDP per capita (current US$; 2023 estimate, World Bank, see note in Table 2 on data sourcing) presented to deal with the confound that AIPI being a composite of infrastructure and human-capital measures is partly a measure of income Shannon et al. The selection of income control (instead of a more extensive set of institutional covariates) reflect the standard augmented-Solow reasoning by (Mankiw et al., 1992); we address the lacking presence with a more complete institutional-quality control in Section 7.

Table 2. Compiled dataset

Country	**AIPI (2023)**	**Real GDP growth, 2023 (%)**	**GDP per capita, current US$***
Singapore	0.80	1.1	84,734
Denmark	0.78	1.9	68,044
United States	0.77	2.5	82,769
United Kingdom	0.73	0.3	47,251
China	0.64	5.2	12,484
Indonesia	0.52	5.0	4,193
India	0.49	7.0	2,271
Sri Lanka	0.43	-2.3	3,828
Bangladesh	0.38	5.8	1,885
Mauritania	0.233	4.8	2,136
Sudan	0.233	-12.0	678
South Sudan	0.11	-8.0	1,072†
Central African Republic	0.18	0.6	397

** GDP per capita, current US$, World Bank national accounts data (series NY.GDP.PCAP.CD), single 2023 vintage for all countries. † South Sudan has no official World Bank estimate more recent than 2015 (the last available figure, used here); the absence of an updated series is itself indicative of the statistical-capacity constraints affecting fragile states discussed in Section 4.1, and the growth figure for South Sudan in this table is compiled from IMF staff estimates rather than a national-accounts GDP-per-capita series that would permit an updated figure. The identical AIPI value for Mauritania and Sudan (0.233) was independently cross-checked against a second source computed directly from the underlying IMF indicator series; both countries genuinely share this value in the primary data, and it is not a transcription artifact.*

Source: Fund (2024); IMF (2023, 2024b); World Bank national accounts data (NY.GDP.PCAP.CD).

Table 3 offers descriptive statistics for all three variables in the total sample and without the two outliers from conflict-affected states identified below in Section 4.3; the decline of standard deviation of growth from 5.57 to 2.86 points once Sudan and South Sudan are excluded anticipates the leverage result formalized in Table 5.

Table 3. Descriptive statistics

Variable	Mean	SD	Min	Max
Panel A. Full sample (N = 13)				
AIPI (0–1 scale)	0.484	0.246	0.110	0.800
Real GDP growth (%)	0.92	5.57	−12.00	7.00
GDP per capita (current US$)	23,980	33,691	397	84,734
Panel B. Excluding conflict outliers (N = 11)				
AIPI (0–1 scale)	0.541	0.221	0.180	0.800
Real GDP growth (%)	2.90	2.86	−2.30	7.00
GDP per capita (current US$)	28,181	35,155	397	84,734

Note: *SD = standard deviation (n−1 divisor). 'Conflict outliers' refers to Sudan and South Sudan (see Section 4.3).*

Source: authors' calculations based on Table 2.

Income confounds shown directly in Fig 3. In this sample, AIPI and log GDP per capita are extremely highly correlated (Pearson r = 0.95), which we think is the empirical analogue of the theoretical concern noted just above we have no cross-sectional design this small that can separate an AI specific channel from a general income/institutional-quality channel with two very collinear regressors. In Section 5 we nevertheless report the income-controlled specification, for full transparency and completeness, but with a cautionary note in Section 4.4 about its limited power.

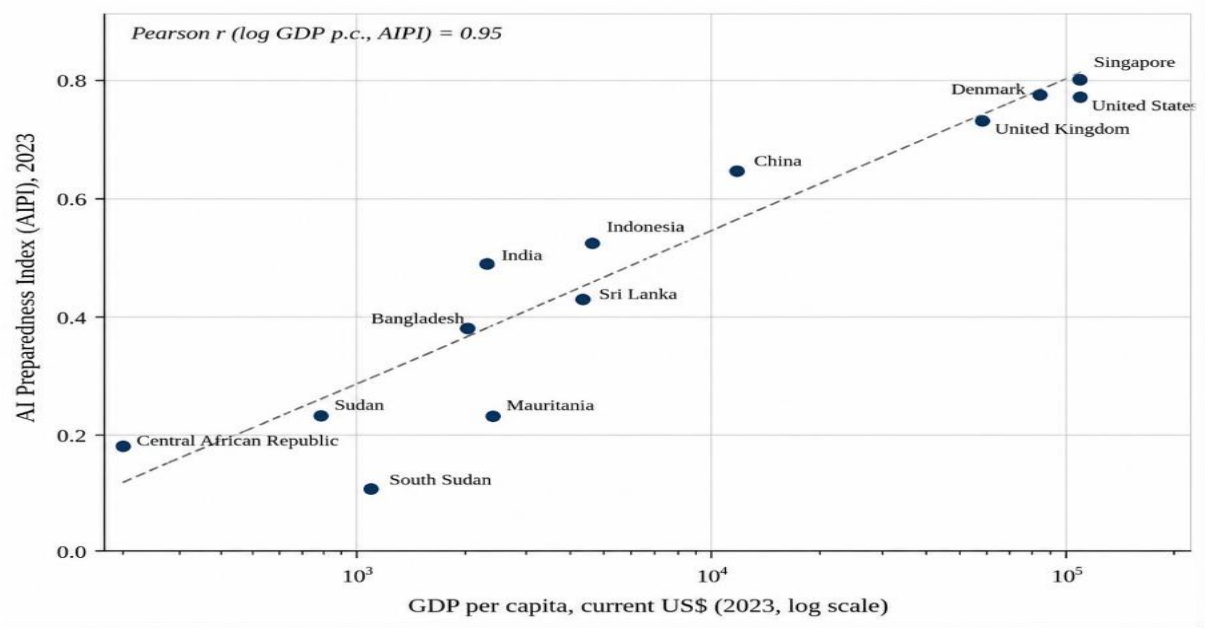


Figure 3. AI Preparedness Index and GDP per capita, 2023 (the income confound)

Note: *x-axis on a log scale; dashed line is a log-linear fitted trend (not a causal estimate).*

Source: authors' chart based on Fund (2024) and World Bank national accounts data (NY.GDP.PCAP.CD), as compiled in Table 2.

Estimation strategy

We run both a bivariate OLS model GDP growth$_i = \beta_0 + \beta_1 \cdot \text{AIPI}_i + \varepsilon_i$ and then add an extension by controlling for log GDP per capita. In all specifications, we report the Pearson coefficient of correlation and R^2, coefficients along with significance at a 0.05 level σ and a 95% confidence for β_1. As robustness checks, we (a) re-estimate excluding Sudan and South Sudan, countries whose 2023 growth collapse is driven by ongoing civil conflict rather than technology; (b) use Spearman's rank correlation as a bivariate nonparametric check, given the sensitivity to outliers already evident in (a); and (c) present Cook's distance statistics (Cook et al., 1977). In light of the small sample, typical OLS standard errors are also the most susceptible to heteroskedasticity of any specification option in this paper; we report the standard (homoscedastic) OLS confidence intervals throughout for comparability with both bivariate and multivariate specifications as well, and flag heteroskedasticity-robust standard errors (White, 1980) as an obvious refinement for larger panels proposed in Section 7 where the distinction is more than a rounding error due to additional degrees of freedom.

Limitations of the design

This design has insufficient statistical power with n = 13 and cannot support conclusions regarding causality. Even the regression added in Section 4.2 does not discriminate between AI-specific effects and those due to general income and institutional-quality which AIPI partly proxies for since a lone cross-sectional year provides so few observations (13) on so many possibly correlated regressors (2) that the multivariate model itself has too weak a power to isolate either effect very precisely (discussed further in Section 5). Reverse causality is also possible: AIPI increases with the infrastructure and education investment to which richer, faster-growing countries can afford. In Section 7 we prescribe the panel design required for that, treating this paper as a purely exploratory, hypothesis generating exercise rather than as a definitive test.

Results

The bivariate regression on the full 13-country sample yields:

$$\text{GDP growth} = -3.66 + 9.44 \cdot \text{AIPI}, \quad R^2 = 0.174, \quad \text{Pearson } r = 0.418, \quad p = 0.156, \quad 95\%\text{ CI for } \beta_1 = [-4.19, 23.07]$$

The coefficient on AIPI is positive at first glance, as the naive "AI readiness leads to growth" story would have it, but it's not statistically significant by normal standards (the 95% confidence interval comfortably contains zero) and the model only accounts for 17% of cross-country variance in 2023 growth rates. The cell-type-specific nonparametric Spearman correlation is even smaller (ρ = 0.264, p = 0.383), indicating that a relationship that is dependent on the exact functional form of the test itself is not robust.

Table 4. Regression results

Specification	**N**	**Pearson r**	**p-value**	**R^2**	**β (AIPI)**	**95% CI (β_1)**
Bivariate, full sample	13	0.418	0.156	0.174	9.44	[−4.19, 23.07]
Bivariate, excl. conflict outliers	11	−0.151	0.657	0.023	−1.96	[−11.62, 7.70]
Multivariate (+ log GDP p.c.), full sample	13	—	0.141[a]	0.274	30.89	[−12.14, 73.93]
Multivariate (+ log GDP p.c.), excl. outliers	11	—	0.362[a]	0.162	14.03	[−19.45, 47.50]

[a] *p-value on the AIPI coefficient in the multivariate specification (log GDP per capita included as a control).*

Source: authors' calculations based on Table 2 (OLS estimation; see Section 4.3 for specification details).

Including the income-level control boosts explanatory power to some degree R^2 increases from 0.174 to 0.274 in the full sample but AIPI stays statistically insignificant ($p = 0.141$) with a 95% CI that comfortably covers zero ([−12.14, 73.93]) and we note that this time around both log GDP-per-capita itself is also negligible and also of negative sign which runs counter to expectation (most probably an example of collinearity between AIPI and income level in a dataset with only thirteen observations, rather than here showing an actual negative effect). Removing the control (N = 11) plus the two outliers with respect to the conflict leaves this same picture: $R^2 = 0.162$, p(AIPI) = 0.362. This is informative in its own right: it indicates that the null AIPI–growth finding does not stem from an omitted income confound masking a real relationship - controlling for income does not unveil some covert growth-enhancing AIPI effect.

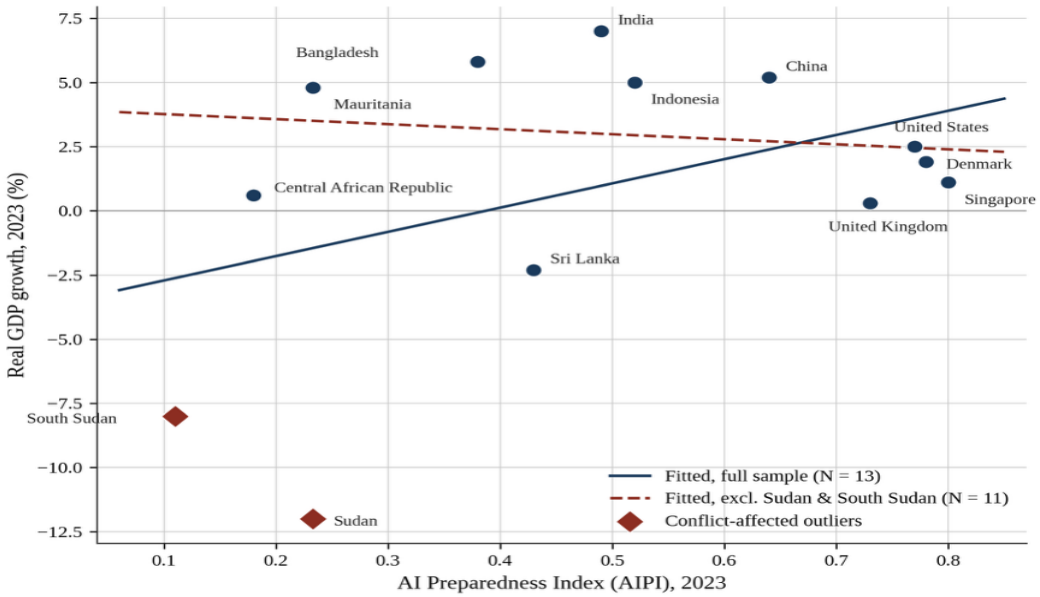


Figure 4. AI Preparedness and Real GDP Growth, 2023 (N = 13)

Note: *solid line is the fitted bivariate regression, full sample (N = 13); dashed line is the fitted regression excluding Sudan and South Sudan (N = 11, diamond markers). The two conflict-affected outliers sit far below the trend implied by the remaining 11 countries and are the leverage points reported in Table 5.*

Source: authors' chart and calculations based on Table 2.

Without Sudan and South Sudan two developing countries in 2023 where the conflicts causing major output collapse are fully outside of the control of AI readiness, this estimated connection is not just purified, but weakly sign-reversed, ($r = −0.151$, $p = 0.657$). According to Table 5, this is not only a matter of visual impression: Sudan's Cook (0.467) distance and South Sudan's (0.262) are by far the highest in the analysis sequence with values from two to seven times those for any other observation indicating that the slight positive correlation in the complete sample is driven almost entirely by these two points rather than reflecting a consistent systematic AIPI-growth relationship among all other non-conflict economies.

Table 5. Case-influence diagnostics (Cook's distance, full-sample bivariate model)

Country	**Cook's distance**
Singapore	0.048
Denmark	0.018
United States	0.006

United Kingdom	0.035
China	0.020
Indonesia	0.023
India	0.059
Sri Lanka	0.013
Bangladesh	0.069
Mauritania	0.164
Sudan	0.467
South Sudan	0.262
Central African Republic	0.038

Note: *Cook's distance ((Cook, 1977)) computed for the full-sample bivariate model reported in Table 4, row 1.*

Source: authors' calculations based on Table 2.

These values are plotted directly in Figure 5, along with the conventional (but often arbitrary) threshold of medium-sized leverage points as approximated by 4 / N ≈ 0.308 which some investigators use again informal but a rough rule-of-thumb for flagging leverage points (Belsley et al., 1980) only (Cook, 1977) for Sudan exceeds this threshold outright at 0.467; South Sudan is only just shy of that threshold at a still-quite-large value of 0.262 or ~33% larger than any other observation).

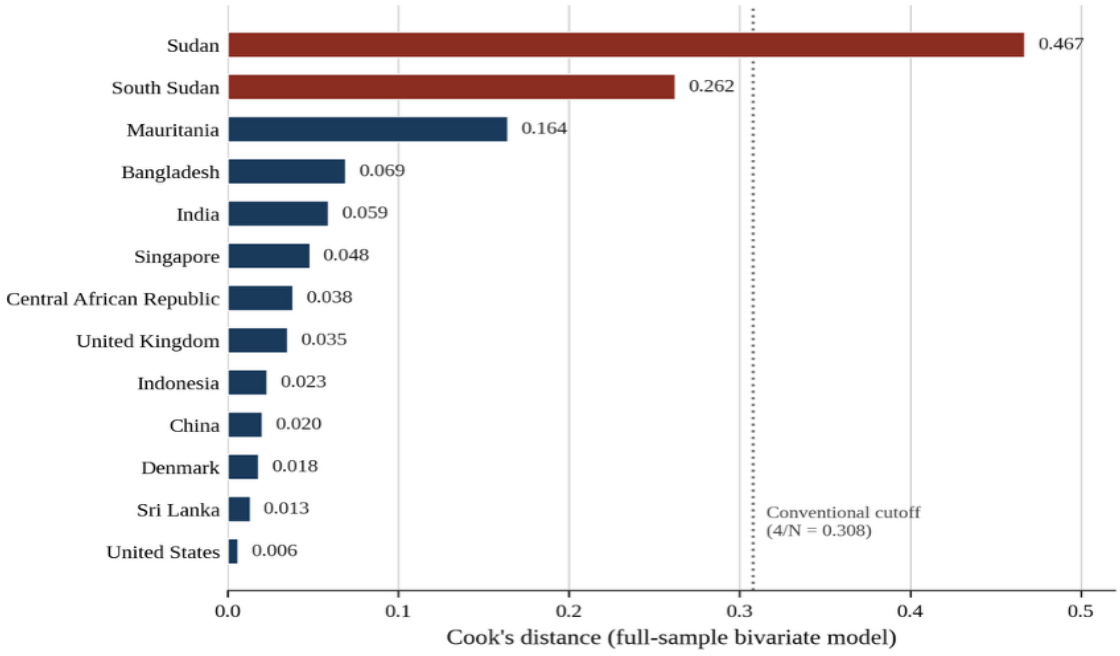


Figure 5. Case-influence diagnostics by country

Note: *dotted vertical line marks the conventional 4/N cutoff (4/13 = 0.308) for flagging high-leverage observations.*

Source: authors' calculations based on Table 2 (see also Table 5).

DISCUSSION

Three findings stand out. For one, in this sample there is no evidence of a significant relationship between similar-year growth in preparedness for AI in the bivariate and income-controlled specifications alike, as well as both parametric tests and nonparametric alternatives which cannot be statistically distinguished from zero.

Secondly, this pattern is logically consistent with the task-based theory of Section 3 and its historical usages in Sections 2.3–2.4. Without going overboard on structural econometrics, (Acemoglu, 2025) own back of the envelope estimate is an useful anchor here: even with quite favorable assumptions about what labels/box-plot tasks AI touches and how much it saves from each, at imaginable values none which give a definitive specification for US TFP effect fail to surpass one percentile points ave. total factor productivity in the US over ten years natural magnitudes that are not detectable via thirteen-country, one-year cross-section using a standard deviation in growth rates of greater than five points unless perfect model-external validity coincides exactly with the Acemoglu estimated relationship. Close to a worst-case design for finding an AI growth relationships, even one that will become economically large (but is not currently reflected in profits) if reinstatement effects take time and complementary investment exactly the mechanism documented for prior general-purpose technologies in the productivity J-curve literature (Brynjolfsson et al., 2021) and now echoed in the firm-level finding that enterprise generative AI deployments mostly fail to show up in profit and loss on average as of late 2025 has been a same-year cross-section. The lack of a contemporaneous relationship is thus evidence about the timing of this relationship, but not evidence against an eventual AI-growth link.

Third, it is consistent with well-established growth economics that some of the fastest-growing economies in the sample India (7.0%), Bangladesh (5.8%), China (5.2%) have no more than low to middle AIPI scores, while some of the highest-AIPI economies (Singapore, United Kingdom) grew comparatively slowly in 2023 despite much greater AI readiness outputs compared to low-AIPI countries (1.1%, 0.3%)(Barro, 1991); (Mankiw et al., 1992) for reasons unrelated to AI but relating instead to conditional convergence and capital deepening effects especially tapping into emerging labor forces far less saturated in terms of capital at levels of human maturity where high productivity occurs across a nation capable of effectively deploying way beyond intensive-output economic capacity). This is exactly the challenge that the income-level control in Section 5 was meant to explore, and its lack of significance suggests a more nuanced story than an income proxy, a question still open for the panel extension discussed below.

This pattern is illustrated in Figure 6. If you rank the sample by AIPI (top panel), then plot the corresponding 2023 growth rate for each set of countries in order (bottom panel), there's hardly a consistent gradient: some of the lowest-AIPI, non-conflict economies such as India and Bangladesh, Mauritania had several of the highest growth rates anywhere on the sample while many high-AIPI economies like Singapore and Denmark and even United States and Canada met with comparatively modest rates. Experiment

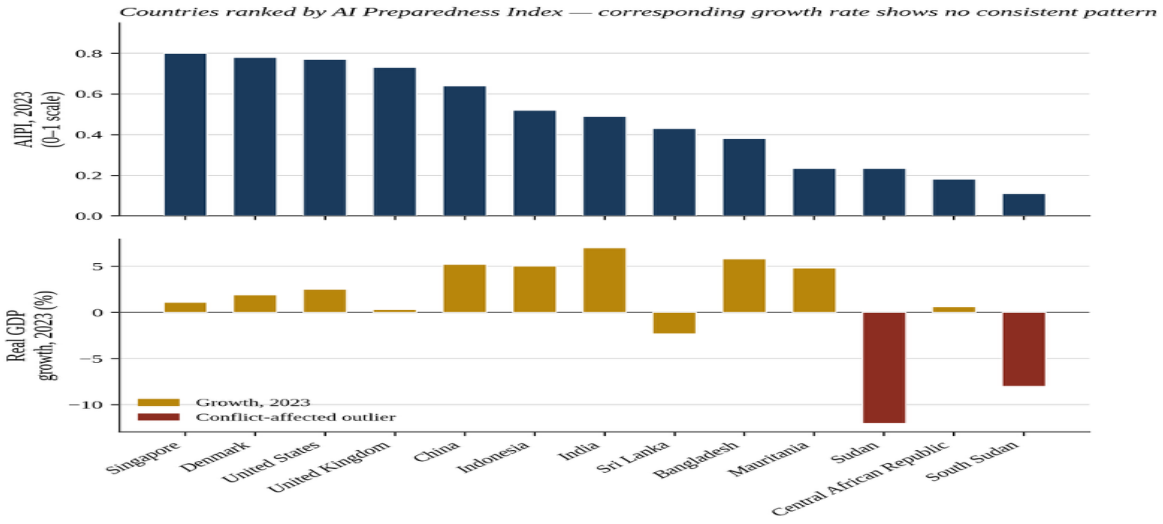


Figure 6. AI Preparedness Index and real GDP growth, ranked by AIPI

Note: *countries ordered by descending AIPI in both panels; the maroon bars in the bottom panel mark the two conflict-affected outliers (Sudan, South Sudan).*

Source: authors' chart based on Table 2.

Limitations and Agenda for Future Research

This paper's principal limitation is sample size and sample construction: n = 13, purposely selected to span the AIPI range. We recommend the following concrete extensions, in priority order, before any stronger empirical claim is advanced:

1. Replicate on the full available cross-section (up to 174 economies) using bulk-downloaded AIPI and WEO data matched by ISO country code and a single reporting vintage, rather than the hand-verified subsample used here.

2. Extend to a multi-year panel (e.g., 2023–2025 as further AIPI releases become available) with country and year fixed effects, which is the design implied by Section 3's own prediction that AI-growth effects should appear with a lag, and which is also the design needed to distinguish a genuine AIPI effect from the kind of diffusion-speed effects documented by (Comin & Hobijn, 2010)

3. Use a smoothed dependent variable (e.g., 3-year average growth) as an additional robustness check against single-year conflict or commodity shocks, which is a less severe alternative to case-deletion for handling the Sudan/South Sudan pattern documented in Section 5.

4. Add institutional-quality controls (e.g., a governance or rule-of-law index) alongside income, since AIPI's regulation and ethics pillar likely correlates with both.

5. When the full-panel replication in (1) is carried out, cross-check every AIPI value directly against the primary IMF dashboard as a matter of course, following the same practice used for the 13-country sample in this paper, and report heteroskedasticity-robust standard errors (White, 1980) alongside the conventional ones given the larger but still cross-sectionally heterogeneous sample.

A cross-sectional design also cannot rule out reverse causality: richer, faster-growing countries can better afford the digital infrastructure and education investments that raise AIPI, so any relationship that does emerge in a larger sample would still require an instrumental-variable or panel fixed-effects strategy before being read causally.

CONCLUSION

This paper provides the first direct test of the relationship between the IMF's AI Preparedness Index and realized 2023 economic growth. Across bivariate and income-controlled specifications, parametric and nonparametric tests, and full-sample and outlier-robust subsamples, we find no statistically significant relationship. We interpret this null result as consistent with task-based automation theory and with the historical pattern of technology-productivity lags documented for earlier general-purpose technologies a pattern now visible again, at the firm level, in the gap between real task-level generative AI productivity gains and their slow appearance in enterprise profit and loss. The central contribution of this paper is not the point estimate itself which a sample of 13 countries cannot support as a definitive finding but a transparent, fully-diagnosed demonstration that a same-year cross-section of AI readiness carries no discernible growth signal, together with a concrete, prioritized agenda (Section 7) for the panel study needed to test whether that signal emerges with the theoretically predicted lag.

Declarations

Funding: No external funding was received for this project.

Conflicts of interest: The author(s) declare no conflict of interest;

Data availability: All data used in this paper are from publicly available sources (fully documented in Table 2 and Appendix A (IMF AI Preparedness Index Dashboard; IMF World Economic Outlook database; World Bank national accounts data). The dataset and the code are available upon reasonable request from the corresponding author.

Author contributions: *[Mahin Mashud]*

REFERENCES

1. Acemoglu, D. (2025). The simple macroeconomics of AI. Economic Policy, 40(121), 13–58. https://doi.org/10.1093/epolic/eiae042
2. Acemoglu, D., & Restrepo, P. (2019). Automation and New Tasks: How Technology Displaces and Reinstates Labor. Journal of Economic Perspectives, 33(2), 3–30. https://doi.org/10.1257/jep.33.2.3
3. Acemoglu, D., & Restrepo, P. (2020). Robots and Jobs: Evidence from US Labor Markets. Journal of Political Economy, 128(6), 2188–2244. https://doi.org/10.1086/705716
4. Aghion, P., J. B. F., & J. C. I. (2019). (2019). Artificial Intelligence and Economic Growth. In The Economics of Artificial Intelligence (pp. 237–290). University of Chicago Press. https://doi.org/10.7208/chicago/9780226613475.003.0009
5. Barro, R. J. (1991). Economic Growth in a Cross Section of Countries. The Quarterly Journal of Economics, 106(2), 407. https://doi.org/10.2307/2937943
6. Belsley, D. A., Kuh, E., & Welsch, R. E. (1980). Regression Diagnostics. Wiley. https://doi.org/10.1002/0471725153
7. Bresnahan, T. F., & Trajtenberg, M. (1995). General purpose technologies 'Engines of growth'? Journal of Econometrics, 65(1), 83–108. https://doi.org/10.1016/0304-4076(94)01598-T
8. Brynjolfsson, E. (1993). The productivity paradox of information technology. Communications of the ACM, 36(12), 66–77. https://doi.org/10.1145/163298.163309
9. Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative AI at Work. The Quarterly Journal of Economics, 140(2), 889–942. https://doi.org/10.1093/qje/qjae044
10. Brynjolfsson, E., Rock, D., & Syverson, C. (2021). The Productivity J-Curve: How Intangibles Complement General Purpose Technologies. American Economic Journal: Macroeconomics, 13(1), 333–372. https://doi.org/10.1257/mac.20180386
11. Cazzaniga, M., Jaumotte, F., Li, L., Melina, G., Panton, A., Pizzinelli, C., Rockall, E., & Mendes Tavares, M. (2024). Gen-AI. Staff Discussion Notes, 2024(001), 1. https://doi.org/10.5089/9798400262548.006
12. Comin, D., & Hobijn, B. (2010). An Exploration of Technology Diffusion. American Economic Review, 100(5), 2031–2059. https://doi.org/10.1257/aer.100.5.2031
13. Cook, R. D. (1977). Detection of Influential Observation in Linear Regression. Technometrics, 19(1), 15–18. https://doi.org/10.1080/00401706.1977.10489493
14. Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of Artificial Intelligence on Knowledge Worker Productivity and Quality. https://doi.org/10.2139/ssrn.4573321
15. Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? Technological Forecasting and Social Change, 114, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019
16. Grossman, G. M., & Rossi-Hansberg, E. (2008). Trading Tasks: A Simple Theory of Offshoring. American Economic Review, 98(5), 1978–1997. https://doi.org/10.1257/aer.98.5.1978
17. Jorgenson, D. W., & Stiroh, K. J. (2000). Raising the Speed Limit: U.S. Economic Growth in the Information Age. Brookings Papers on Economic Activity, 2000(1), 125–210. https://doi.org/10.1353/eca.2000.0008
18. Mankiw, N. G., Romer, D., & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth. The Quarterly Journal of Economics, 107(2), 407–437. https://doi.org/10.2307/2118477
19. Noy, S., & Zhang, W. (2023). Experimental evidence on the productivity effects of generative artificial intelligence. Science, 381(6654), 187–192. https://doi.org/10.1126/science.adh2586

20. Ponce Del Castillo, A. (2024). Generative AI, generating precariousness for workers? AI & SOCIETY, 39(5), 2601–2602. https://doi.org/10.1007/s00146-023-01719-9
21. Solow, R. M. (1956). A Contribution to the Theory of Economic Growth. The Quarterly Journal of Economics, 70(1), 65. https://doi.org/10.2307/1884513
22. White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48(4), 817. https://doi.org/10.2307/1912934
23. Zeira, J. (1998). Workers, Machines, and Economic Growth. The Quarterly Journal of Economics, 113(4), 1091–1117. https://doi.org/10.1162/003355398555847
24. International Monetary Fund. (2023). World Economic Outlook database, October 2023. IMF.
25. International Monetary Fund. (2024a). AI Preparedness Index Dashboard: Technical note. IMF, Washington, DC. International Monetary Fund. (2024b). World Economic Outlook database, October 2024. IMF
26. Levine, R., & Renelt, D. (1992). A Sensitivity Analysis of Cross-Country Growth Regressions. The American Economic Review, 82(4), 942–963. http://www.jstor.org/stable/2117352
27. MIT NANDA. (2025). The GenAI divide: State of AI in business 2025. MIT Media Lab, Massachusetts Institute of Technology.
28. Oxford Insights. (2023). Government AI Readiness Index 2023.

Appendix A. Data Sources and Provenance

Table A1 documents, variable by variable, the primary source, vintage, and access notes underlying the compiled dataset in Table 2, to support independent verification and replication.

Table A1. Data sources and provenance

Variable	Definition	Primary source	Vintage / notes
AIPI	Composite AI-readiness score (0–1) across four pillars: digital infrastructure; human capital & labor-market policy; innovation & economic integration; regulation & ethics	IMF AI Preparedness Index Dashboard (Fund, 2024)	2023 data, released June 2024; cross-checked against a second, independently computed source for every sample country
Real GDP growth	Annual % change, constant prices	IMF World Economic Outlook database (Fund, 2024; (IMF), 2023)	2023; fragile-state figures (Sudan, South Sudan) compiled from IMF staff country reporting where WEO coverage is limited
GDP per capita	Current US$ (used only as a log-transformed control variable)	World Bank national accounts data, series NY.GDP.PCAP.CD	2023 (South Sudan: 2015, the latest available official estimate)

Note: *this table complements, rather than duplicates, the per-table source notes given throughout the paper (Tables 2–5); it is provided as a single consolidated reference for replication.*

Source: authors' compilation.