DISCUSSION PAPER SERIES

DP19883

COMPARATIVE ADVANTAGE IN AI-INTENSIVE INDUSTRIES: EVIDENCE FROM US IMPORTS

Alessandra Bonfiglioli, Rosario Crinò, Mattia Filomena and Gino Gancia

INTERNATIONAL TRADE AND REGIONAL ECONOMICS AND MACROECONOMICS AND GROWTH

ISSN 0265-8003

COMPARATIVE ADVANTAGE IN AI-INTENSIVE INDUSTRIES: EVIDENCE FROM US IMPORTS

Alessandra Bonfiglioli, Rosario Crinò, Mattia Filomena and Gino Gancia

Discussion Paper DP19883
Published 27 January 2025
Submitted 15 January 2025

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
187 boulevard Saint-Germain, 75007 Paris, France
Tel: +44 (0)20 7183 8801
www.cepr.org

COMPARATIVE ADVANTAGE IN AI-INTENSIVE INDUSTRIES: EVIDENCE FROM US IMPORTS

Abstract

This paper investigates the determinants of comparative advantage in Artificial Intelligence (AI)-intensive industries using a comprehensive dataset of US imports from 68 countries across 79 manufacturing and service industries over the period 1999–2019. Using a novel measure of AI intensity based on the prevalence of occupations requiring expertise in machine learning and data analysis, we identify key factors influencing exports in AI-intensive industries. Our analysis reveals that countries with larger STEM graduate populations, broader Internet penetration and higher export volumes exhibit stronger export performance in AI-intensive industries. In contrast, regulatory barriers to digital trade are associated with lower AI-intensive exports. These results are robust to controlling for traditional sources of comparative advantage and addressing potential threats to identification. Our findings have implications for understanding competitiveness in the digital economy and highlight that fostering capabilities in data-driven industries may be particularly important due to their pronounced scale economies.

JEL Classification: F10, F14, J23

Keywords: N/A

Alessandra Bonfiglioli - alessandra.bonfiglioli@unibg.it
University of Bergamo and CEPR

Rosario Crinò - rosario.crino@unibg.it
University of Bergamo and CESifo and CEPR

Mattia Filomena - mattia.filomena@unibg.it
University of Bergamo

Gino Gancia - ginoalessandro.gancia@unimib.it
University of Milano-Bicocca and CEPR

Acknowledgements
We thank Anna Salomons and seminar participants at the 1st CEPR Workshop of the Research and Policy Network on Artificial Intelligence (Paris, 2024), the Marche Polytechnic University and the University of Bergamo for useful comments. Rosario Crinò gratefully acknowledges financial support from the Italian Ministry for University and Research under the PRIN 2022 program (project title: AUTOmation, PROductivity and Wage INequality (AUTOPROWIN): Firms and Workers in Times of Economic Turmoil; project number: 2022FLBY7J; CUP: F53D23003060006), M4, C2, Investment 1.1 - financed by the European Union -- Next Generation EU, D.D. MUR 104, February 2, 2022.

Comparative Advantage in AI-Intensive Industries: Evidence from US Imports

Alessandra Bonfiglioli[a,*], Rosario Crinò[b], Mattia Filomena[c] and Gino Gancia[d]

[a] *University of Bergamo and CEPR*

[b] *University of Bergamo, CEPR and CESifo*

[c] *University of Bergamo*

[d] *University of Milano-Bicocca and CEPR*

January 15, 2025

Abstract

This paper investigates the determinants of comparative advantage in Artificial Intelligence (AI)-intensive industries using a comprehensive dataset of US imports from 68 countries across 79 manufacturing and service industries over the period 1999–2019. Using a novel measure of AI intensity based on the prevalence of occupations requiring expertise in machine learning and data analysis, we identify key factors influencing exports in AI-intensive industries. Our analysis reveals that countries with larger STEM graduate populations, broader Internet penetration and higher export volumes exhibit stronger export performance in AI-intensive industries. In contrast, regulatory barriers to digital trade are associated with lower AI-intensive exports. These results are robust to controlling for traditional sources of comparative advantage and addressing potential threats to identification. Our findings have implications for understanding competitiveness in the digital economy and highlight that fostering capabilities in data-driven industries may be particularly important due to their pronounced scale economies.

Keywords: Artificial Intelligence, international trade, digital data, comparative advantage.

JEL Classification: F10, F14, J23.

---

[*]Corresponding author. Department of Economics, University of Bergamo, via dei Caniana 2, 24127, Bergamo, Italy. E-mail addresses: alessandra.bonfiglioli@unibg.it (A. Bonfiglioli), rosario.crino@unibg.it (R. Crinò), mattia.filomena@unibg.it (M. Filomena), ginoalessandro.gancia@unimib.it (G. Gancia). We thank Anna Salomons and seminar participants at the 1st CEPR Workshop of the Research and Policy Network on Artificial Intelligence (Paris, 2024), the Marche Polytechnic University and the University of Bergamo for useful comments. Rosario Crinò gratefully acknowledges financial support from the Italian Ministry for University and Research under the PRIN 2022 program (project title: AUTOmation, PROductivity and Wage INequality (AUTOPROWIN): Firms and Workers in Times of Economic Turmoil; project number: 2022FLBY7J; CUP: F53D23003060006), M4, C2, Investment 1.1 - financed by the European Union – Next Generation EU, D.D. MUR 104, February 2, 2022.

1 Introduction

Artificial Intelligence (AI) and machine learning are among the most significant driving forces powering manufacturing and service industries today. Their increasing adoption is driven by the need to efficiently manage a growing volume of data. Applications include predictive maintenance, quality control, customization, supply chain management, inventory optimization, pricing algorithms and customer service. AI and machine learning, coupled with big data, enable companies to enhance and automate decision-making processes, which are often crucial for survival, particularly in industries where competition operates on a global scale.

Although AI and machine learning are being widely implemented, their adoption varies significantly across sectors and countries. To deploy AI systems, businesses require a workforce equipped with specific skills, high-quality IT capital and easy access to vast amounts of digital data. These prerequisites are not always readily available. Shortages of qualified workers, inadequate digital infrastructure and a limited customer base can pose critical challenges, especially in industries where AI holds the greatest potential.

In this paper, we study the determinants of Comparative Advantage (CA) in AI-intensive industries. We analyze data on US imports from up to 68 countries across 79 4-digit manufacturing and service industries over the past two decades (1999–2019). Our sample captures the rise of the digital economy. Although machine learning was first introduced in the 1950s, it began to gain traction in the business world in the late 1990s thanks to the increasing availability of digital data from the Internet, advancements in storage technologies and research breakthroughs spearheaded by tech giants. While trade is more prevalent in the manufacturing sector, our sample also includes service industries where AI has numerous direct applications.

To guide the empirical analysis, we begin by presenting a simple neoclassical model of trade, building on Eaton and Kortum (2002), Chor (2010) and Costinot et al. (2012). This framework allows us to derive a testable equation showing that countries export relatively more in industries that intensively use the factors they possess in greater abundance (Heckscher-Ohlin CA) and in industries where they exhibit relatively higher productivity (Ricardian CA). To test these predictions empirically, we combine various proxies for factors facilitating AI adoption at the country level (i.e., potential sources of CA) with data on AI intensity at the industry level.

We treat AI intensity as a technological characteristic of industries and follow Bonfiglioli et al. (2025) in measuring it as the relative employment of AI-related workers in the US. AI-related workers are defined as those employed in occupations that typically require knowledge of specialized software for machine learning and data analysis, as identified in the "Hot Technologies" section of the O*NET database. The top service industries for AI intensity include information and data processing, financial services and other business services. The top manufacturing industries include the production of communication equipment, audio and video equipment, electronic products, navigational and control instruments, and computer equipment.

We consider several sources of CA. First, to test for Heckscher-Ohlin CA, we focus on production factors that are critical in AI-intensive industries. The appropriate type of human capital is probably the most important input, as data analysts and engineers design and adjust the software that powers the algorithms. Therefore, our preferred proxy is the supply of bachelor's, master's and Ph.D. graduates in STEM disciplines from the OECD. As an alternative proxy, we also consider the share of capital in communication and computer equipment relative to total assets, sourced from the EU KLEMS database.

Second, to test for Ricardian CA, we consider a number of factors that may enhance productivity in AI-intensive industries. We use the share of the population with Internet access as a measure of a country's digital infrastructure. Additionally, economies of scale are argued to be particularly significant as a source of competitiveness in data-intensive industries. There are several reasons for this. First, AI-driven processes can be scaled up much more rapidly than traditional processes. Second, a larger customer base generates more data, which can be aggregated across networks to make AI systems even more effective. Therefore, we use the total volume of exports as a proxy for market scale. Intuitively, countries with higher export volumes should be able to collect more customer data, which is a valuable asset in data-intensive industries.

Finally, we also consider some policy measures. Although policy is clearly endogenous and may depend on trade itself, including it in the analysis is an interesting exercise. Given the synergies between algorithms and data flows, and following Sun and Trefler (2023), we use the Digital Services Trade Restrictiveness Index (DSTRI) and its components. These indexes, developed by the OECD, measure various dimensions of regulatory barriers in digital services trade and data flows across countries (Ferencz, 2019).

We find significant empirical support for all the main sources of CA: greater availability of STEM graduates or ICT capital, broader Internet diffusion and higher volumes of total exports are associated with increased exports to the US in AI-intensive industries. In contrast, stricter regulatory barriers to digital trade are associated with lower AI-intensive exports. These results are robust to the inclusion of standard sources of CA considered in the literature and a host of fixed effects, as well as to the use of alternative proxies and econometric specifications to control for sample selection and confounding factors. By exploiting predetermined variation in the regressors and a historical instrument for STEM graduates, we also show that our results are unlikely to be driven by reverse causality.

Quantitatively, we find that all factors driving CA in AI-intensive industries are roughly equally important in explaining differences in export volumes across countries and industries. For example, a larger population share of new STEM graduates, equivalent to the average difference between France and Mexico, is associated with 0.27 log points higher exports to the US in "Other General Purpose Machinery Manufacturing" (high AI intensity) compared to "Metal Forging and Stamping" (low AI intensity). Since 1999, however, the main driver of CA has been digital infrastructure, due to its greater time variation. Consistent with this result, regulations on infrastructure and connectivity are the most detrimental policy restrictions, followed by a residual category that includes performance requirements, limitations on downloading and streaming, and restrictions on online advertising, among other measures.

Our results have clear policy relevance, as they identify factors and policy bottlenecks crucial for competitiveness in AI-intensive industries at the global level. At the same time, they provide insights into why building CA in these industries may be particularly important. Specifically, our data show that AI-intensive industries have experienced faster export growth than the average industry, indicating that they are among the most dynamic sectors in the global economy. Furthermore, these industries are characterized by stronger scale economies compared to more traditional sectors, suggesting that CA may become self-reinforcing and that specialization in these sectors could yield greater economic gains.

Our analysis relates to the empirical literature investigating the role of sources of CA, such as differences in factor endowments (Romalis, 2004), institutions (Nunn, 2007; Levchenko, 2007) and financial development (Manova, 2013; Bonfiglioli et al., 2019). It also connects to the empirical literature based on multisector extensions of the Ricardian model in Eaton and Kortum (2002), especially Chor (2010) and Costinot et al. (2012). We differ by focusing specifically on the determinants of CA in AI-intensive industries. To the best of our knowledge, this is the first paper to do so.

Our paper is also related, albeit less closely, to the literature on the effects of digital technologies on trade. Goldfarb and Trefler (2019) and Ferencz et al. (2022) provide qualitative discussions of how AI may impact international trade. Empirically, Sun and Trefler (2022) and Sun and Trefler (2023) study international trade in mobile app services. Using an instrument inspired by the Heckscher-Ohlin model, they find that the use of AI increases app downloads and expands their user base in foreign countries. Brynjolfsson et al. (2019) show that eBay's introduction of a machine translation system significantly increased its exports.

There is also a large literature on the effect of the Internet on trade. Papers such as Freund and Weinhold (2004), Blum and Goldfarb (2006), Lendle et al. (2016), Chen and Wu (2021), and Carballo et al. (2022) show that e-commerce websites facilitate exports. Bailey et al. (2021) demonstrate that social connections between countries, derived from Facebook data, are strong determinants of bilateral trade flows, especially for goods where information frictions are significant. Demir et al. (2025) document that improved Internet infrastructure enhances supply chain efficiency and expands firms' access to suppliers, using firm-to-firm transaction data for Turkey. Rather than studying how AI and digital technologies facilitate trade, this paper explores the determinants of exports in AI-intensive industries.

2 Theoretical Framework

To guide the empirical analysis, in this section, we describe a neoclassical model of trade and CA (both Heckscher-Ohlin and Ricardian) based on Eaton and Kortum (2002), Chor (2010) and Costinot et al. (2012). The goal is to illustrate the main sources of CA in AI-intensive industries and derive estimation equations that can be used to test them.

Consider a world with $n = 1, ..., N$ countries and $i = 1, ..., I$ industries. In each country, consumers have a two-tier utility function. The upper tier utility function is Cobb-Douglas, while the lower tier is constant elasticity of substitution:

$$U_n = \exp\left[\sum_i \ln\left(\int_0^1 \left(Q_n^i(j)\right)^\sigma \mathrm{d}j\right)^{\alpha^i/\sigma}\right], \tag{1}$$

where $Q_n^i(j)$ denotes the quantity of variety j from industry i consumed in country n and $\sum_i \alpha^i = 1$. The parameter $\epsilon = 1/(1-\sigma) > 1$ is the elasticity of substitution between any two varieties from the same industry. Demand for a variety with price $p_n^i(j)$ is:

$$Q_n^i(j) = \left[p_n^i(j)/p_n^i\right]^{1-\sigma} \alpha^i E_n, \tag{2}$$

where E_n is total expenditure in country n and $p_n^i = \left[\sum_i p_n^i(j)^{1-\sigma}\right]^{1/(1-\sigma)}$.

Let $p_{no}^i(j)$ denote the price that country o would charge for exporting variety j to country n. Due to perfect competition, $p_{no}^i(j)$ is equal to the marginal cost in origin o:

$$p_{no}^i(j) = \frac{c_o^i d_{no}^i}{z_o^i(j)}, \tag{3}$$

where c_o^i is the unit production cost of the prospective exporter o in industry i, while $d_{no}^i \geq 1$ is the iceberg transport cost (with $d_{nn}^i = 1$ and $d_{no}^i \leq d_{nk}^i d_{io}^i$) and $z_o^i(j)$ captures the Ricardian productivity of country o in manufacturing variety j. The unit production cost c_o^i is a Cobb-Douglas aggregate over factor prices in country o:

$$c_o^i = \prod_{f=0}^{F} (w_{of})^{s_f^i}, \tag{4}$$

where w_{of} is the local unit price of factor f and s_f^i is its cost share, with $\sum_f s_f^i = 1$.

As in Eaton and Kortum (2002), $z_o^i(j)$ is drawn from a Frechét distribution with cdf $F_o^i(z) = \exp\left[-\left(z/A_o^i\right)^{-\theta}\right]$. The parameter $A_o^i > 0$ controls for the average level of technology of country o in industry i, and $\theta > 1$ is an inverse measure of dispersion. Under these assumptions, the probability that o exports any of an industry-i variety to country n is:

$$\pi_{no}^i = \frac{\left[A_o^i / \left(c_o^i d_{no}^i\right)\right]^\theta}{\sum_{o=1}^N \left[A_o^i / \left(c_o^i d_{no}^i\right)\right]^\theta}. \tag{5}$$

As in Eaton and Kortum (2002), country n's average expenditure per good does not vary by source. Hence, the fraction of goods that country n buys from country o in industry i, π_{no}^i, is also the fraction of expenditure on imports from o. Denote X_{no}^i the value of industry-i imports from country o. Trade flows normalized by expenditure on locally-sourced goods, X_{nn}^i, are:

$$\frac{X_{no}^i}{X_{nn}^i} = \frac{\pi_{no}^i}{\pi_{nn}^i} = \frac{\left[A_o^i / \left(c_o^i d_{no}^i\right)\right]^\theta}{\left[A_n^i / c_n^i\right]^\theta}. \tag{6}$$

Note that X_{no}^i / X_{nn}^i depends on bilateral variables only.

Finally, to close the model, we impose market clearing for each factor f and country o:

$$\sum_{i=1}^I \sum_{n=1}^N s_f^i X_{no}^i = w_{of} V_{of}, \tag{7}$$

where V_{of} is country o's endowment of factor f. This system of equations cannot be solved analytically. However, because of transport costs, there will be in general no factor price equalization, implying that relatively more abundant factors will have relatively lower factor prices.

We now derive the main equation to be estimated. To this end, we first specify the cost of trade as the product of the following components:

$$d_{no}^i = \tau_i \cdot \tau_{no} \cdot \tau_{no}^i. \tag{8}$$

That is, d_{no}^i depends on an industry component τ_i reflecting technological characteristics such as the bulk weigh of products in an industry, a country-pair component τ_{no} reflecting geographic characteristics such as distance, and finally a country-pair&industy component τ^i_{no} such as trade policy. Next, using c^i_i, d^i_{no} and $s^i_0 = 1 - \sum_{f=1}^{F} s^i_f$ into (6) and taking logs, we can write the equation:

$$\ln X^i_{no} = -\theta \sum_{f=1}^{F} \left(\ln \frac{w_{of}}{w_{o0}} \right) s^i_f + \theta \ln \tau^i_{no} + \theta \ln A^i_o + \delta_o + \delta_{ni}, \tag{9}$$

where δ_o is an exporter fixed effect and δ_{ni} an importer-industry fixed effect. Notice that the fixed effects absorb all the terms in the denominator of (6) and also all the technological and geographical components of the cost of trade.

Since factor prices are endogenous, we follow the approach taken by Romalis (2004) and treat relative factor prices as an inverse function of relative factor endowments.[1] Moreover, since we will use data for a single destination country (the United States), we remove the destination index n:

$$\ln X^i_o = \theta \sum_{f=1}^{F} \left(\ln \frac{V_{of}}{V_{o0}} \right) s^i_f + \theta \ln A^i_o + \theta \ln \tau^i_o + \delta_o + \delta_i. \tag{10}$$

The first term in (10) captures the Heckscher-Ohlin forces: in general, country o exports relatively more in industries that use more intensively the relatively abundant factors. In the specific case of AI-related exports, this prediction will be tested with an interaction between an industry-level measure of AI intensity and the country-level endowment of scientific skills.

The second term in (10) captures other technological determinants of CA. For AI-intensive industries and besides the effect through the supply of scientific skills, the literature has emphasized some other important factors (see, for instance, Goldfarb and Trefler, 2019). First, the development of a country's digital infrastructure is certainly a key determinant of productivity in AI-intensive industries. Second, it has also been argued that economies of scale are especially important for the adoption of AI and other digital technologies. One of the main reasons is that more customers generate more data, which increases the value of investing in data analysis. In turn, data analysis can improve decision making, which allows firms to reach even more customers. To capture these factors, we allow A^i_o to depend on a linear combination of the AI intensity of the industry and some relevant country characteristics:

$$\ln A^i_o = \alpha_I \cdot s^i_{AI} \cdot I_o + \alpha_X \cdot s^i_{AI} \cdot X_o, \tag{11}$$

where I_o is a proxy for country's o digital infrastructure and $X_o = \sum_{n=1}^{N} \sum_{i=1}^{I} X^i_{no}$ is the total volume of country's o exports, a proxy for the overall level of demand. Positive estimates of the parameters

[1] If relative factor endowments are a noisy proxy for relative factor prices, it would bias the results against finding significant effects.

α_I and α_X will indicate that AI-intensive exports are more sensitive than the rest of exports to digital infrastructure and scale, respectively.[2]

3 Data and Preliminary Evidence

In this section, we describe the data and the main variables used in the empirical analysis. Then, we present preliminary evidence on CA and trade in AI-intensive industries.

3.1 Data and Variables

Our analysis combines information on three characteristics of countries and industries: (1) exports to a single destination market at the country-industry level; (2) AI intensity of production in each industry; and (3) country characteristics that may give rise to CA in AI-intensive industries. In this section, we present the main variables and data sources. Appendix A provides additional details, as well as information on the other variables used in the analysis. The sample encompasses 79 industries covering the entire manufacturing and service sectors. The number of countries varies from 45 to 68 depending on the source of CA considered in the analysis.[3] The period of analysis spans from 1999 to 2019, with the starting year determined by the availability of trade data (which are missing for the service industries prior to 1999) and the end year chosen to exclude the COVID-19 pandemic.

3.1.1 Exports

We focus on the US as the destination of countries' exports. Besides being the main foreign market for most countries in the world, the US offers high-quality data on bilateral trade for a long time period and across a large number of industries, both in manufacturing and in services. We obtain the trade data from two sources. For the manufacturing sector, we use bilateral data on US imports from Feenstra et al. (2002), covering virtually all countries in the world. These data are recorded according to the Harmonized System (HS) classification and provide a mapping between HS 10-digit product codes and industry codes, according to the North American Industry Classification System (NAICS). We aggregate the product-level US import data from each origin country and year at the 4-digit level

[2]In a similar vein, Chor (2010) also allows productivity, A_o^i, to be affected by other linear combinations of country characteristics—namely, financial development and institutional quality—and industry characteristics—namely, financial dependence and contract intensity. We will consider these variables as well as other traditional determinants of CA as additional controls.

[3]The full list of countries is provided in Appendix Table A.1.

of NAICS, in order to be consistent with the industry detail of the AI intensity measure (described below). The number of manufacturing industries is 66. For the service sector, we exploit bilateral data on US imports from the Bureau of Economic Analysis (BEA). These data are available for 13 service industries and 71 origin countries from the year 1999. Given our focus on exporting countries, from now on we use “exports to the US”, or simply “exports”, when referring to the US import data.

3.1.2 AI Intensity of Industries

Industries that rely more heavily on AI in their production processes use more extensively a variety of advanced software to handle, organize and process vast amounts of digital data, as well as to design, customize and implement machine learning algorithms. As a result, these industries also have a greater need for workers with advanced knowledge of specialized software compared to other industries. Such an advanced knowledge is primarily possessed by professionals engaged in a narrow range of occupations related to the domains of computer science, mathematics, statistics and operations research. To measure the AI intensity of each industry, therefore, we use information on the relative employment of a set of AI-related occupations, which we identify based on the specialized knowledge of data processing and machine learning software required for workers in each job.

Following Bonfiglioli et al. (2025), we exploit a novel section of the O*NET database called “Hot Technologies”. The latter reports the software requirements that are most frequently included in all current employer job postings in the US, separately for each occupation in the 2018 Standard Occupational Classification (SOC). The original software list contains 157 titles, which span from software with general applications like Microsoft Excel, to advanced programming languages like Python and C++. With the help of computer scientists, we identify 54 software (see Appendix Table A.2) that are normally used for data collection and generation, execution and adaptation of machine learning algorithms, and to feed these algorithms with large datasets.

With the list of specialized software at hand, we use the “Hot Technology” section of O*NET to identify occupations for which each software is “in demand”. These comprise 82 occupations whose job postings typically require knowledge of a given software. We refine this list through two sequential filters. First, we exclude occupations for which only one software is “in demand”. This filter excludes 21 occupations that use a single software in their daily activities, like “Special Effects Artists and Animators”, who only use Python, or “Commercial and Industrial Designers”, who only use JavaScript. Second, we select occupations that require skills within the typical domains of AI. To this purpose, we combine occupational definitions from the SOC with the official crosswalk between the SOC and the Classification of Instructional Programs (CIP). Using this information, we identify

Table 1: AI-Related Occupations

SOC Code	SOC Definition	SOC Code	SOC Definition
151211	Computer Systems Analysts	151253	Software Quality Assurance Analysts and Testers
151221	Computer and Information Research Scientists	151254	Web Developers
151231	Computer Network Support Specialists	151255	Web and Digital Interface Designers
151232	Computer User Support Specialists	151299	Computer Occupations, All Others
151241	Computer Network Architects	152021	Mathematicians
151242	Database Administrators	152031	Operations Research Analysts
151243	Database Architects	152041	Statisticians
151244	Network and Computer Systems Administrators	152051	Data Scientists
151251	Computer Programmers	439111	Statistical Assistants
151252	Software Developers		

Notes: Occupations are classified according to the 6-digit level of the 2018 Standard Occupational Classification (SOC).

occupations with non administrative, supportive or educational roles, whose required skills are provided by academic programs in the following fields: computer and information sciences and support services, mathematics and statistics, and operations research.[4] This filter eliminates occupations like "Economists". While both Python and SQL are "in demand" for this job, the main skills required for this occupation are in the realm of social sciences according to the SOC-CIP crosswalk. The final list of AI-related occupations comprises 19 titles, which are listed in Table 1 with the corresponding 6-digit SOC codes.

To measure the AI intensity of each industry, we combine the list of AI-related occupations with micro-level employment data from the 2020 American Community Survey (ACS), using information on each worker's SOC occupation and NAICS industry available in the ACS.[5] Then, we measure the AI intensity of each industry i as the ratio of employment in AI-related occupations (L_i^{AI}) to employment in all other—non-AI-related—occupations (L_i^{nAI}) in the industry:

$$AIint_i = \frac{L_i^{AI}}{L_i^{nAI}}. \tag{12}$$

This measure is available for 79 4-digit NAICS industries, which correspond to those for which we have export data as described above.

[4]The crosswalk between SOC and CIP is provided by the National Center for Education Statistics of the US Department of Education. The corresponding CIP codes among the occupations that pass the first filter are: 110101, 110201, 110701, 111006, 143701, 270503 and 520302.

[5]The ACS is a 1% random sample of the US population. To increase sample size, we follow Autor et al. (2013) and Acemoglu and Restrepo (2020) by using pooled 5-year ACS data for 2017-2021. This approach also ensures that the 2020 employment figures are comparable to those for the years 1980 and 2000, which will be used later in the analysis and are based on 5% random samples of the US Census of Population.

Table 2: Top and Bottom Industries in Terms of AI Intensity

Top Ten Industries	$AIint_i$	Bottom Ten Industries	$AIint_i$
Information and Data Processing Services	0.2633	Dairy Products	0.0089
Computer Equipment Manufacturing	0.2190	Seafood and Other Miscellaneous Foods, nec	0.0088
Communications, Audio and Video Equipment	0.1415	Sawmills and Wood Preservation	0.0087
Other Business Services	0.1363	Miscellaneous Nonmetallic Mineral Products	0.0086
Telecommunication Services	0.1305	Other Woods Products	0.0085
Electronic Components and Products	0.1015	Cement, Concrete, Lime, and Gypsum Products	0.0074
Navigational Electronic and Control Instruments Manufacturing	0.1000	Animal Slaughtering and Processing	0.0070
Financial Services	0.0953	Bakeries	0.0031
Audiovisual Services	0.0915	Construction	0.0030
Aerospace Products Manufacturing	0.0871	Fiber, Yarn, and Thread Mills	0.0028

Notes: $AIint_i$ is the ratio of employment in AI-related occupations to employment in non-AI-related occupations in each industry in 2020. Industries are classified according to the NAICS 4-digit classification.

$AIint_i$ is equal to 0.034 on average and exhibits significant variation across industries, with a standard deviation of 0.046. Table 2 reports the value of $AIint_i$ in the top and bottom 10 industries in our sample. In line with Bonfiglioli et al. (2025), the top 10 industries consist of advanced services such as information and data processing services, other business services, telecommunication services, financial services and audiovisual services. They also include advanced branches of the manufacturing sector producing computer equipment, communication, audio and video equipment, electronic components and control instruments, and aerospace products. On the contrary, the bottom 10 industries largely consist of traditional manufacturing activities, such as bakeries, animal processing, construction, wood products and nonmetallic mineral products.[6]

Over time, the ranking of industries in terms of AI intensity has remained largely stable: the correlation between the 2000 and 2020 rankings exceeds 0.88.[7] This high stability reflects the fact that the ranking captures technological differences across industries that persist over time, rather than transitory fluctuations or time-varying shocks to the values of $AIint_i$. By construction, moreover, the ranking cannot expand or shrink over time. For these reasons, and because our primary interest lies in the relative position of industries rather than the actual values of their AI intensity, we use the ranking of industries by $AIint_i$ as our baseline measure. We rank industries in ascending order of $AIint_i$, so higher positions in the ranking correspond to higher levels of AI intensity. In Section 6.2, we show that our main conclusions are robust across alternative measures of AI intensity, including the direct use of $AIint_i$.

[6]The full list of industries with their AI intensity levels is presented in Appendix Table B.1.

[7]We measure AI intensity in 2000 as in (12), using micro-level employment data from the US Census of Population. To track AI-related occupations back in time across the revisions of the SOC that occurred between 2000 and 2020, we use correspondence tables from the US Bureau of Labor Statistics.

3.1.3 Sources of Comparative Advantage

The first country characteristic we examine as a source of CA in AI-intensive industries is the availability of advanced scientific skills. The high requirements of these industries in terms of data collection and implementation of machine learning algorithms make scientific skills a critical input to production. We measure the availability of scientific skills in a given country c and year t using information on new graduates in STEM fields (science, technology, engineering and mathematics) from the OECD Education Statistics "Graduates by Fields" database. Our proxy is defined as the number of new bachelor's, master's and Ph.D. graduates in STEM fields per 100 inhabitants aged 15-64 (S_{ct}). This variable is available over the period 1999-2019 for 45 of the 71 countries for which we also have export data. These countries belong to four geographical areas: Europe and Central Asia, East Asia and Pacific, America and the Caribbean, Middle East and South Africa.

As shown in Table 3, S_{ct} equals 0.24 on average and varies from 0.15 in "America and the Caribbean" to 0.29 in "East Asia and Pacific". S_{ct} has significantly increased between 1999 and 2019, growing by 0.12 on average across countries. The number of new STEM graduates has grown not only relative to the population but also relative to the total number of graduates, increasing from 22% to 24% on average over the sample period. This suggests that the composition of countries' graduate populations has progressively shifted toward scientific degrees.

There are two main caveats to using S_{ct} as a source of CA in AI-intensive industries. First, not all STEM degrees are likely to equip individuals with skills that are equally relevant to these industries. While we use the total number of new STEM graduates throughout the analysis—primarily because this variable is not subject to excessive transitory fluctuations in most countries—we dig deeper into the role played by different degrees in Section 6.2, using the number of new STEM graduates by field. As expected, we find that CA is primarily driven by the availability of skills from digitally-oriented fields.

Second, S_{ct} is intended to capture the availability of the most advanced scientific skills in each country, as these skills are particularly valuable for AI-intensive industries. However, not all new STEM graduates possess such advanced skills. Some may have intermediate-level skills that are unlikely to significantly influence CA in AI-intensive industries. This suggests that S_{ct} may be subject to measurement error, which could result in a downward bias in its coefficient. Although the available data do not provide information on the specific skill levels of individuals, in Section 7.2, we present evidence of this downward bias using an Instrumental Variables approach based on historical data on the number of famous scientists in each country.

We consider two additional sources of CA that are expected to enhance the productivity of AI-

Table 3: Sources of Comparative Advantage: Descriptive Statistics

	New STEM Graduates per 100 Inhabitants (S_{ct})			Population Share with Internet Access (I_{ct})			Total Exports (X_{ct})		
	Mean	SD	Δ Mean	Mean	SD	Δ Mean	Mean	SD	Δ Mean
Total	0.237	0.101	0.120	0.491	0.295	0.676	210,049.1	320,351.7	160,966.1
Europe and Central Asia	0.244	0.102	0.138	0.589	0.264	0.703	222,774.2	282,176.5	153,351.0
East Asia and Pacific	0.286	0.082	0.082	0.464	0.304	0.610	374,806.2	488,004.6	357,710.8
America and Caribbean	0.154	0.058	0.058	0.348	0.269	0.617	90,850.1	140,504.6	48,876.3
Africa and Middle East	0.232	0.083	0.083	0.399	0.299	0.812	68,869.8	73,649.8	42,480.8

Notes: The subscripts c and t denote countries and years, respectively. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15–64. I_{ct} is the share of the population with Internet access. X_{ct} is total exports in millions of US dollars. The means and standard deviations are computed across all countries within a given geographical area and across all years. Average changes are calculated as the average difference in a given variable between the first and the last sample year across all countries within a given region.

intensive industries. First, we account for cross-country heterogeneity in digital infrastructure by using the share of the population with Internet access in each country and year (I_{ct}), as reported in the World Bank World Development Indicators (WDI) database. While this variable may not fully capture the most advanced forms of digital investment, its overall variation likely reflects key differences in digital infrastructure across countries. Additionally, this variable has broad coverage, being available for nearly all countries in our sample—68 out of the 71 countries with export data. As shown in Table 3, I_{ct} averages 49% but varies significantly across regions, ranging from 35% in "America and the Caribbean" to 59% in "Europe and Central Asia". Internet coverage has also grown substantially over time, both on average and in each region.

Finally, we examine the role of economies of scale, which can influence the competitiveness of AI-intensive industries for two main reasons. On the one hand, AI-driven processes can be scaled up more rapidly than traditional processes. On the other hand, a larger number of consumers generate more data, which enhances the performance of AI systems. To capture this source of CA, we use data on each country's total exports (X_{ct}), sourced from the WDI and available for the same 68 countries with data on Internet access. Intuitively, countries with higher export volumes should be able to collect data from more customers, a particularly valuable asset for AI-intensive industries.[8] As with the other proxies, Table 3 shows that X_{ct} varies significantly across regions and has increased substantially over the sample period.

[8] Our main results remain robust when using exports to the US instead of total exports, although the former is more likely to be mechanically related to our dependent variable.

3.2 Preliminary Evidence

We now present some new facts about trade in AI-intensive industries and its relationship with the sources of CA. We start by studying how countries' exports to the US have changed over the sample period in AI-intensive industries relative to other industries. To this purpose, we estimate the following specification:

$$\Delta \ln M_{ci} = \alpha_c + \lambda \mathbb{I}_i + \varepsilon_{ci}, \tag{13}$$

where $\Delta \ln M_{ci}$ is the log change in US imports from country c in industry i between the first and the last sample year; $\mathbb{I}_i$ is a dummy equal to 1 for AI-intensive industries and 0 otherwise; α_c are country fixed effects; and ε_{ci} is an error term. We estimate this specification on the sample of 68 countries. The coefficient λ measures the average difference in the growth rate of exports to the US in AI-intensive industries relative to other industries within countries.

The estimates of λ are reported in Table 4, where we use three progressively more restrictive definitions of AI-intensive industries. In column (1), industries are classified as AI intensive if their $AIint_i$ values are above the sample median, while industries with below-median $AIint_i$ are classified as non AI intensive. In column (2), we compare industries in the top tercile of $AIint_i$ (AI intensive) to industries in the bottom tercile (non AI intensive). Finally, in column (3), AI-intensive industries are defined as those in the top quintile of $AIint_i$, while non-AI-intensive industries are those in the bottom quintile. The coefficient λ is consistently positive and very precisely estimated, increasing in size as the definition of AI-intensive industries becomes more stringent. These results indicate that exports have been growing significantly faster in AI-intensive industries relative to other industries over the sample period. The difference is as large as 0.24 log points under the most conservative definition and increases to 0.53 log points when comparing the top and bottom quintiles.

We now provide an initial overview of how CA in AI-intensive industries varies across countries. To achieve this, we compare countries based on a measure of revealed CA in these industries, defined as follows:

$$RCA_{ct} = \frac{\frac{\sum_{i \in AI} M_{cit}}{\sum_i M_{cit}}}{\frac{\sum_c \sum_{i \in AI} M_{cit}}{\sum_c \sum_i M_{cit}}}, \tag{14}$$

where AI denotes the set of AI-intensive industries and M_{cit} are US imports from country c in industry i and year t. RCA_{ct} is akin to the Balassa (1965) index but evaluates revealed CA of each country c,

Table 4: Export Growth: AI-Intensive vs. Other Industries

	Above vs. Below Median (1)	Top vs. Bottom Terciles (2)	Top vs. Bottom Quintiles (3)
AI-Intensive ($\mathbb{I}_i$)	0.235***	0.409***	0.528***
	(0.048)	(0.070)	(0.090)
Country FE	Yes	Yes	Yes
Observations	5,203	3,466	2,061
Adj. R^2	0.157	0.149	0.167

Notes: The table reports the estimates of (13). The dependent variable is the log change in exports to the US from country c in industry i between the first and the last sample year. $\mathbb{I}_i$ is a dummy equal to 1 for AI-intensive industries and 0 otherwise. AI-intensive (non-AI-intensive) industries are those with $AIint_i$ above (below) the sample median in column (1), in the top (bottom) tercile of the distribution in column (2) and in the top (bottom) quintile of the distribution in column (3). Standard errors, reported in parentheses, are corrected for clustering within countries. * Significant at 10%; ** significant at 5%; *** significant at 1%.

relative to all other countries in the sample, within the US market rather than the world as a whole.[9] A value of RCA_{ct} above 1 indicates that country c has a revealed CA in AI-intensive industries in the US market.

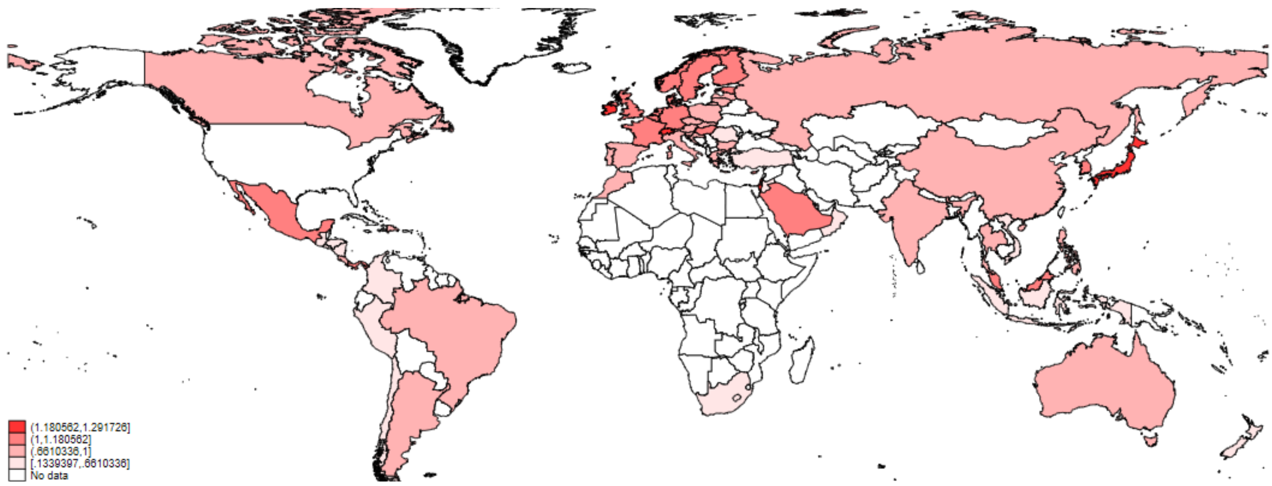
Figure 1 plots the average values of RCA_{ct} over the sample period. AI-intensive industries are defined as those with $AIint_i$ above the sample median. The figure unveils significant variation in revealed CA across countries. The most technologically advanced economies, such as Ireland, Switzerland, Japan, the UK and Northern European countries stand out for their high values of RCA_{ct}. The index also exceeds 1 in countries like Mexico, Saudi Arabia and Malaysia, reflecting their specialization in industries such as automotive, chemicals, petroleum refining, insurance and other business services, which exhibit relatively high levels of AI intensity. In contrast, the lowest values of RCA_{ct} are observed in Latin America and Southeast Asia.

We conclude this section with preliminary evidence on the relationship between countries' exports and the sources of CA. To this end, we first compute total exports to the US from each country c and year t, separately in two groups of industries with $AIint_i$ above the median (AI intensive) and below the median (non AI intensive). Then, we regress the log of these exports on a full set of country fixed effects to neutralize differences in geographical distance and other country-level determinants of trade. Finally, we compute the average residuals from these regressions for ten groups of countries, corresponding to the deciles of the average value of a given determinant of CA over the sample period.

Figure 2, panels (a)-(c), plot these quantities against the ten deciles of S_{ct}, I_{ct} and X_{ct}, separately for AI-intensive and non-AI-intensive industries. The plots reveal that in AI-intensive industries,

[9] Country-level exports to the world are unavailable for the service industries included in our sample.

Figure 1: Revealed Comparative Advantage in AI-Intensive Industries



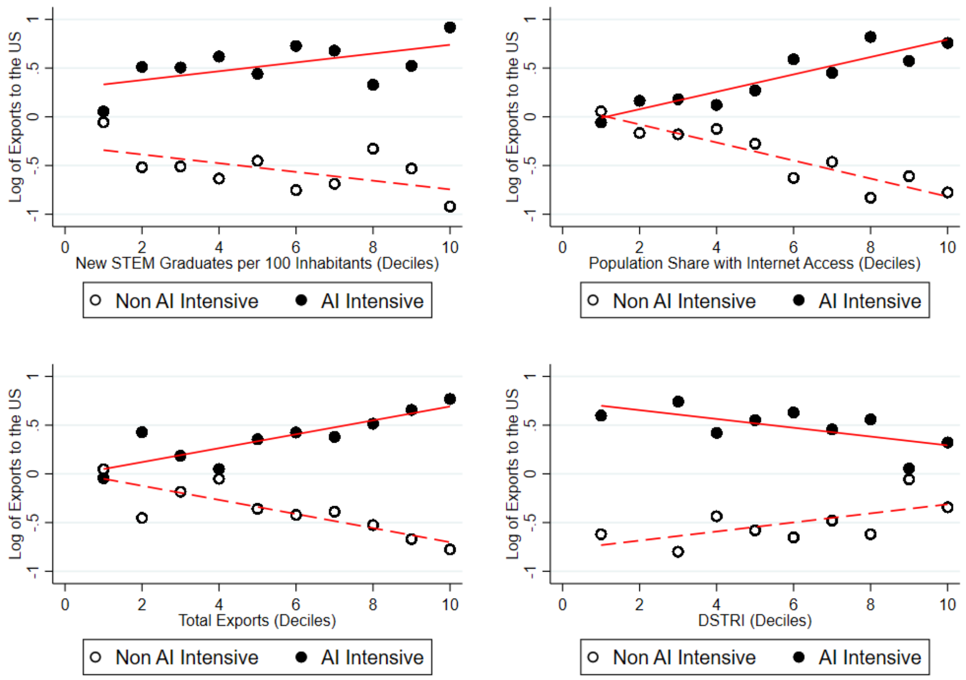
Notes: The map reports the average values of the revealed comparative advantage index in (14). Averages are computed over all sample years.

countries' exports increase sharply with each of the three sources of CA, while in other industries, they slightly decrease. Hence, countries with greater availability of scientific skills, more advanced digital infrastructure and stronger economies of scale export relatively more to the US in industries with a higher AI intensity, consistent with these countries enjoying a CA in these industries.

Finally, panel (d) presents similar evidence for a policy-based measure of restrictions on digitally-enabled trade: the Digital Services Trade Restrictiveness Index (DSTRI) from the OECD. This composite index measures the restrictiveness of regulations in five areas: (1) infrastructure and connectivity, (2) electronic transactions, (3) e-payment systems, (4) intellectual property rights and (5) other barriers to trade in digitally-enabled services. The index ranges from 0 to 1, where 0 indicates a fully open regulatory environment and 1 represents a completely closed regime. The index is available for the period 2014–2019 and covers 53 of the 71 countries with export data.

The figure indicates that exports decrease across the deciles of DSTRI for AI-intensive industries, while this is not the case for other industries. Thus, countries with a more liberal environment for digitally-enabled trade (lower deciles of DSTRI) export relatively more to the US in AI-intensive industries. This finding suggests that specific aspects of regulation may serve as additional sources of CA beyond those stemming from country characteristics. In the following sections, we systematically analyze the roles of scientific skills, digital infrastructure and economies of scale. We return to the role of regulation in Section 8.

Figure 2: Exports, AI Intensity and Country Characteristics



Notes: The figure shows the average residuals obtained by regressing the log of exports to the US from country c in industry i and year t on country fixed effects. Averages are computed separately for AI-intensive and non-AI-intensive industries across ten deciles of the distribution of the country characteristic indicated on the horizontal axis of each graph. AI-intensive industries are those with $AIint_i$ above the sample median. Linear regression lines are shown as solid lines (AI-intensive industries) or dashed lines (non-AI-intensive industries). DSTRI is an index of the restrictiveness of regulations on digitally-enabled trade. Higher values correspond to a more regulated environment.

4 Empirical Specification and Identification Strategy

Following the theoretical model (eq. 10 and 11), we estimate variants of the following specification:

$$\ln M_{cit} = \alpha_{ct} + \alpha_{it} + \beta\left(Z_{ct} \times AIint_i^r\right) + \boldsymbol{\Gamma}_{cit}\boldsymbol{\gamma} + \varepsilon_{cit}, \tag{15}$$

where M_{cit} denotes US imports from country c in industry i and year t; Z_{ct} is one (or a combination) of the sources of CA for country c in year t; $AIint_i^r$ is the ranking of industries by AI intensity; $\boldsymbol{\Gamma}_{cit}$ is a vector of control variables specific to each country c, industry i and year t; and ε_{cit} is an error term. We account for two sets of fixed effects. First, α_{it} are industry-year fixed effects, which capture heterogeneous trends across industries and industry-specific shocks that affect exports to the US from all origin countries; these include both demand and technology shocks. Second, α_{ct} are country-year fixed effects, which control for heterogeneous trends across countries and any country characteristics influencing exports to the US across all industries. These fixed effects also absorb all bilateral determinants of trade, including those that vary over time.[10]

The coefficient of interest is β, which measures the differential relation between a given source of CA and exports across industries with varying levels of AI intensity. In particular, $\beta > 0$ implies that countries with a higher value of a certain characteristic (scientific skills, digital infrastructure and economies of scale) export relatively more to the US in industries with higher AI intensity, which is interpreted as evidence of CA following a long empirical literature initiated by Romalis (2004).

Since the specification in (15) includes industry-year and country-year fixed effects, the identification of β rests on the interplay between cross-industry variation in AI intensity and cross-country variation in CA within a given year. This empirical strategy mitigates concerns about confounding factors, because potential time-varying confounders operating either at the industry or at the country level are accounted for by the two sets of fixed effects.

Nevertheless, identification could be threatened in two scenarios. The first concern is the potential presence of a confounder that (i) is correlated with the sources of CA and (ii) exerts heterogeneous effects on exports across industries. In Section 7.1, we employ various approaches to accommodate such potential confounders and find no significant changes in the main results. The second concern relates to reverse causality, i.e., the possibility that a shock induces changes in exports in a specific country and industry, which in turn affects the main explanatory variables. In Section 7.2, we propose two approaches that leverage predetermined variation in the regressors and find that the main evidence remains robust.

5 Baseline Results

The baseline estimates of (15) are presented in Table 5. Standard errors, shown in parentheses, are corrected for two-way clustering within country-industry and industry-year pairs; this accounts for residual correlation both over time within each country and industry, and across countries within each industry and year.[11] In column (1), we begin with a parsimonious specification that includes only the interaction between the share of new STEM graduates in the population (S_{ct}) and $AIint_i^r$, along with country-year and industry-year fixed effects. The coefficient β is positive and very precisely

[10]The country-year and industry-year fixed effects also subsume the linear terms in Z_{ct} and $AIint_i^r$.

[11]As shown in Appendix Figure B.1, the main conclusions remain unchanged across several alternative clustering schemes. These include clustering at the industry level as well as two-way clustering by industry and country, by industry and country-year, by industry-year and country, by country-industry and country-year, and by country-year and industry-year.

Table 5: Baseline Estimates

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$S_{ct} \times AIint_i^r$	0.073***		0.094***			0.057**	0.051**
	(0.014)		(0.022)			(0.022)	(0.022)
$I_{ct} \times AIint_i^r$				0.063***		0.040***	0.027***
				(0.005)		(0.008)	(0.009)
$X_{ct} \times AIint_i^r$					0.008***	0.004***	0.004***
					(0.001)	(0.001)	(0.001)
$G_{ct} \times AIint_i^r$		0.012***	-0.006			-0.004	-0.004
		(0.003)	(0.005)			(0.005)	(0.005)
Controls	No	No	No	No	No	No	Yes
Country-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49,812	49,887	49,742	93,531	93,458	49,535	49,535
Adj. R^2	0.658	0.657	0.658	0.629	0.625	0.662	0.663

Notes: The dependent variable is the log of exports to the US from country c in industry i and year t. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. Controls include the interactions of skill and capital endowment of each country c in year t with the ranking of industries in terms of skill and capital intensity in 2020. Standard errors, reported in parenthesis, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

estimated, implying that countries with a larger share of new STEM graduates in their population export relatively more to the US in industries with higher AI intensity.

This result suggests that the availability of advanced scientific skills is a source of CA in AI-intensive industries. An alternative interpretation, however, is that CA in these industries depends on the availability of general skills, with S_{ct} simply serving as a proxy for them. To investigate this possibility, in column (2), we replace the interaction between S_{ct} and $AIint_i^r$ with an analogous interaction using the share of total new graduates in the population (G_{ct}) instead of the share of new STEM graduates. In column (3), we include this new variable alongside the primary interaction between S_{ct} and $AIint_i^r$.

The coefficient of the interaction term between total graduates and AI intensity is positive and highly significant when this variable is included on its own (column 2). However, it turns negative and imprecisely estimated when included together with the STEM graduates-AI intensity interaction (column 3). At the same time, the coefficient on the latter interaction remains positive, highly precise and comparable in magnitude to the estimate reported in column (1). These findings are inconsistent with the view that CA in AI-intensive industries relies on advanced skills in general. Instead, they support the view that CA in these industries is specifically driven by the availability of advanced scientific skills. Going forward, we will always control for the total graduates-AI intensity interaction whenever we include the STEM graduates-AI intensity interaction to disentangle the role of scientific skills from that of general skills.

In columns (4) and (5), we examine the other two determinants of CA. The interactions between the share of the population with Internet access (I_{ct}) and log total exports (X_{ct}) with $AIint_i^r$ both yield positive and highly significant coefficients, indicating that better digital infrastructure and stronger economies of scale are associated with relatively larger exports to the US in AI-intensive industries. In column (6), we include the interactions between $AIint_i^r$ and the three sources of CA jointly. The coefficients retain the same signs as before, remain statistically significant, and their magnitudes are largely stable. This suggests that scientific skills, digital infrastructure and economies of scale are three independent sources of CA in AI-intensive industries.

Finally, in column (7), we extend the specification by adding interactions between proxies for countries' endowments of human and physical capital and proxies for industries' intensities in these production factors (coefficients unreported).[12] This is our preferred specification. The coefficients of interest remain largely unchanged, suggesting that the previous results are not confounded by traditional determinants of CA.[13]

Quantitatively, the estimates in column (7) imply that a larger population share of new STEM graduates, equivalent to the average difference between the country at the 75th percentile of the distribution of S_{ct} (France) and the country at the 25th percentile (Mexico), is associated with exports to the US being 0.27 log points larger in the industry at the 75th percentile of the distribution of $AIint_i$ ("Other General Purpose Machinery Manufacturing") compared to the industry at the 25th percentile ("Metal Forging and Stamping"). A similar difference in the share of the population with Internet access and in total exports is associated with larger exports to the US in the former industry relative to the latter by 0.32 and 0.26 log points, respectively. These magnitudes are comparable to those of a commensurate difference in human capital.

Over the sample period, S_{ct}, I_{ct} and X_{ct} grew, on average, by 0.12, 0.69 and 0.81, respectively. According to the estimates in column (7), this growth is associated with a 0.23, 0.71 and 0.12 log points larger increase in exports in the industry at the 75th percentile of the distribution of $AIint_i$ relative to the industry at the 25th percentile. Overall, these numbers suggest that the role played by the three sources of CA is not only statistically significant but also economically meaningful.

[12] Following Romalis (2004) and Bonfiglioli et al. (2019), we measure each country's skill endowment using the log number of years of schooling and the capital endowment using the log real capital stock per person engaged. These two measures are interacted with industry i's rankings by skill and capital intensity, respectively. Skill intensity is defined as the log ratio between the number of employees with at least a bachelor's degree and those with less than a bachelor's degree. Capital intensity is defined as the log ratio between capital compensation and labor compensation. Both measures are calculated for the year 2020 using US data.

[13] The estimates would be very similar to those in column (7) if the three sources of CA were included one at a time in this specification. See Appendix Table B.2.

6 Sensitivity Analysis

We now conduct an extensive sensitivity analysis to assess the robustness of the baseline results. First, we employ various approaches to address potential influential observations. Second, we examine different measures of AI intensity and alternative proxies for CA. Finally, we use alternative model specifications and estimators.

6.1 Outliers

In Table 6, we show that the baseline results are unlikely to be driven by influential observations. In column (1), we exclude the top and bottom five industries in terms of $AIint_i$. The coefficients are virtually unchanged, suggesting that our evidence is not solely explained by industries at the extremes of the AI intensity distribution. In column (2), we perform a complementary exercise by excluding four industries that are likely involved in the development of AI applications.[14] The coefficients remain unaffected, indicating that the results reflect differences in the intensity of AI adoption rather than being driven by a few industries that develop AI applications.

In columns (3)–(9), we drop countries that fall within the top or bottom 5% tails of various characteristics: the three sources of CA, real GDP (a proxy for country size), real per-capita GDP (a proxy for countries' level of development), growth in import penetration (a proxy for changes in countries' exposure to foreign competition) and the stock of US foreign direct investment over GDP (a proxy for the relative importance of US multinational firms in a given country). The coefficients are remarkably stable across the board, suggesting that our evidence is not driven by a handful of economies with extreme characteristics. Finally, columns (10) and (11) show that the evidence holds when restricting the sample to countries that export to the US in all industries or to industries in which all countries export to the US.

6.2 Alternative Measures

We now consider alternative ways of measuring industries' AI intensity. In column (1) of Table 7, we use the log of $AIint_i$ in place of the ranking $AIint_i^r$ to construct the interactions with the three determinants of CA. Our evidence is preserved. Quantitatively, the estimates imply that a difference in S_{ct}, I_{ct} and X_{ct} corresponding to the interquartile range of the distribution is associated with exports to the US being larger in the industry at the 75th percentile of the distribution of $AIint_i$, relative to

[14] The industries are "Computer Equipment Manufacturing", "Information and Data Processing Services", "Communications, Audio, and Video Equipment" and "Telecommunication Services".

Table 6: Outliers

	AI Intensity		Country Characteristics							Balanced Samples	
	Top & Bottom Five industries (1)	AI Producers (2)	STEM Graduates (3)	Internet Coverage (4)	Total Exports (5)	Real GDP (6)	Real Per Capita GDP (7)	Import Penetration (8)	US FDI (9)	Countries with all Industries (10)	Industries with all Countries (11)
$S_{ct} \times AIint_i^r$	0.052**	0.053**	0.043*	0.056**	0.063**	0.057**	0.057**	0.047**	0.047*	0.064**	0.052*
	(0.026)	(0.024)	(0.025)	(0.025)	(0.026)	(0.024)	(0.022)	(0.023)	(0.026)	(0.026)	(0.028)
$I_{ct} \times AIint_i^r$	0.030***	0.029***	0.021**	0.031***	0.023**	0.021**	0.025***	0.034***	0.023**	0.025**	0.053**
	(0.011)	(0.010)	(0.009)	(0.011)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.010)	(0.020)
$X_{ct} \times AIint_i^r$	0.006***	0.005***	0.004***	0.004***	0.005***	0.004**	0.004***	0.005***	0.005***	0.003*	0.002
	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
$G_{ct} \times AIint_i^r$	-0.002	-0.005	0.002	-0.003	-0.007	-0.003	-0.004	-0.011**	-0.001	-0.005	-0.001
	(0.006)	(0.006)	(0.007)	(0.006)	(0.006)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)	(0.007)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	44,035	47,319	45,076	42,945	43,250	44,119	44,983	43,984	40,253	20,145	6,840
Adj. R^2	0.666	0.664	0.668	0.654	0.634	0.642	0.671	0.673	0.658	0.696	0.725

Notes: The dependent variable is the log of exports to the US from country c in industry i and year t. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. Columns (1) and (2) exclude, respectively, the top and bottom five industries by AI intensity, and industries likely involved in the development of AI applications (i.e., "Computer Equipment Manufacturing", "Information and Data Processing Services", "Communications, Audio, and Video Equipment" and "Telecommunication Services). Columns (3)-(9) exclude countries in the top or bottom 5% of the distribution of the characteristic indicated in each column's heading. Column (10) restricts the sample to countries that have non-missing data on exports to the US in all industries at least once over the sample period. Column (11) restricts the sample to industries for which US imports are recorded from all countries at least once over the sample period. Controls include the interactions of skill and capital endowment of each country c in year t with the ranking of industries in terms of skill and capital intensity in 2020. Standard errors, reported in parenthesis, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

the industry at the 25th percentile, by 0.13, 0.50 and 0.06 log points, respectively, in line with the baseline results.

Next, we use three different classifications of AI-related occupations to reconstruct $AIint_i^r$. In column (2), we define as AI-related all STEM occupations, using the STEM occupational classification of Hanson and Slaughter (2018). In column (3), we narrow down the list by excluding life sciences, medical sciences and other STEM occupations (e.g., surveyors, cartographers and mapping scientists), whose tasks are arguably less AI-intensive than the rest of STEM occupations. Finally, in column (4), we define as AI-related those occupations that fall in the top quartile of the AI-exposure measure proposed by Webb (2020).[15] The coefficients are similar to the baseline estimates, suggesting that our evidence is not an artifact of the definition of AI-related occupations and the resulting measure of industry AI intensity.

Next, we delve deeper into the role of scientific skills. In addition to the total number of new STEM graduates, the OECD provides separate information for four main STEM fields: Information and Communication Technology (ICT); Engineering, Manufacturing and Construction; Natural Sciences; and Mathematics and Statistics. In columns (5)–(8), we estimate (15) by replacing the in-

[15] Webb (2020) identifies patents related to AI and develops an algorithm to count occurrences of similar verb-noun pairs in the patents' titles and in the task descriptions of each occupation. The indicator of AI exposure for each occupation is the average of these common occurrences across the occupation's tasks, weighted by the importance of each task for the occupation.

Table 7: Alternative Measures

	AI Intensity				STEM Fields				Digital Skills
	$AIint_i$	STEM Occupations (All)	STEM Occupations (No Life/Medical/Other)	Webb's Occupational AI Exposure	Natutal Sciences	Mathematics & Statistics	Engineering	ICT	ICT Share of Capital Stock
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$S_{ct} \times AIint_i^r$	1.1050*	0.044**	0.047**	0.081***	-0.036	0.205*	0.046*	0.188***	0.254**
	(0.537)	(0.022)	(0.022)	(0.022)	(0.039)	(0.118)	(0.028)	(0.056)	(0.124)
$I_{ct} \times AIint_i^r$	0.687***	0.039***	0.038***	0.016*	0.031***	0.029***	0.027***	0.030***	0.046***
	(0.203)	(0.008)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)	(0.009)	(0.011)
$X_{ct} \times AIint_i^r$	0.068**	0.007***	0.007***	0.004***	0.005***	0.004***	0.004***	0.004***	0.004***
	(0.029)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$G_{ct} \times AIint_i^r$	-0.039	-0.008	-0.008	-0.019***	0.004	0.003	0.001	-0.004	
	(0.126)	(0.005)	(0.005)	(0.005)	(0.003)	(0.003)	(0.004)	(0.004)	
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49,535	49,535	49,535	49,535	40,706	46,487	43,946	47,114	39,964
Adj. R^2	0.663	0.667	0.667	0.663	0.655	0.655	0.662	0.656	0.691

Notes: The dependent variable is the log of exports to the US from country c in industry i and year t. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. In column (1), $AIint_i^r$ is replaced by $AIint_i$, the log ratio of employment in AI-related occupations to employment in non-AI-related occupations in each industry i. AI-related occupations are defined as in Table 1 in column (1) and in columns (5)-(9); as STEM occupations (Hanson and Slaughter, 2018) in column (2); as STEM occupations excluding life and medical scientists and other STEM occupations in column (3); and as occupations in the top quartile of the AI exposure index proposed by Webb (2020) in column (4). Columns (5)-(8) replace the number of new STEM graduates (S_{ct}) with the number of new graduates in each STEM field, as indicated in each column's heading. The ICT share of capital stock used in place of S_{ct} in column (9) is the share of communication and computer equipment relative to total assets. Controls include the interactions of skill and capital endowment of each country c in year t with the skill and capital intensity of industry i in 2020; skill and capital intensities are the logs of the corresponding proxies in column (1) and the rankings in the other columns. Standard errors, reported in parenthesis, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

teraction between $AIint_i^r$ and the population share of new STEM graduates with the corresponding interactions using the population share of new graduates in each STEM field.

These specifications allow us to examine the distinct roles played by different types of scientific skills. At the same time, the coefficients are likely to be noisier than the baseline estimates, as the number of new graduates by field are more susceptible to transitory fluctuations compared to the total number of new STEM graduates. Nevertheless, consistent with the view that AI-intensive industries are particularly sensitive to scientific skills of a digital nature, the interaction coefficient is large for Mathematics and ICT graduates, while it is much smaller and not always significant for the other categories.

Finally, in column (9), we use the ICT share of capital stock as an alternative proxy for scientific skills. The ICT share of capital stock is highly correlated with the share of new STEM graduates in the population, reflecting the strong complementarity between scientific skills and advanced types of capital.[16] At the same time, this variable provides a stock-based measure of CA. Information on the share of ICT capital is available from the EU KLEMS database for 27 European countries and Japan.

Reassuringly, the main results are unchanged. Quantitatively, a higher ICT share of capital stock, equivalent to the average difference between Sweden (75th percentile) and Malta (25th percentile), is

[16] A regression of S_{ct} on the ICT share of capital stock, controlling for country and year fixed effects, yields a coefficient of 0.026 (s.e. 0.008), with an R^2 of 0.8.

associated with 0.09 log points higher exports to the US in the industry at the 75th percentile of the distribution of $AIint_i$ compared to the industry at the 25th percentile. This magnitude is in the same ballpark as that of a comparable difference in S_{ct}, as discussed in Section 5.

6.3 Alternative Estimators and Specifications

A well-known issue with the estimation of specifications like (15) is that, due to the log transformation, the coefficients are estimated on the (potentially selected) sample of observations with positive trade flows. This issue should not be particularly relevant in our case, as the fraction of country-industry-year triplets with zero exports to the US is relatively small in our sample (around 3.2%), primarily because we focus on a single destination country that represents the main market for most exporters in the world. Nevertheless, in Table 8, we perform a series of alternative exercises to account for possible sample selection.

In column (1), we estimate (15) using the inverse hyperbolic sine transformation, which preserves the zeros, in place of the log transformation. The coefficients are similar to the baseline estimates. In column (2), we implement a two-step approach à la Heckman (1979). We first estimate a Probit model for the probability of observing a triplet with positive exports to the US. Then, we construct the inverse Mills ratio using the predicted values from this regression and include this variable as an additional control in (15). The coefficient on the Mills ratio is 0.896, indicating that the errors of the two equations are correlated. Yet, correcting for sample selection leaves our coefficients of interest close to the baseline estimates.[17] Finally, in column (3), we use the Poisson pseudo-maximum likelihood estimator proposed by Santos Silva and Tenreyro (2006). Except for the coefficient on the interaction between I_{ct} and $AIint_i^r$, which loses significance, the other results are confirmed.

The specification in (15) defines a linear relationship describing how the influence of a given source of CA varies with $AIint_i^r$. To examine the sensitivity of the evidence to the functional form, we now turn to non-parametric results. We begin by estimating the following specification separately for each of the 79 industries:

$$\ln M_{cit} = \alpha_c + \alpha_t + \beta_i^S S_{ct} + \beta_i^I I_{ct} + \beta_i^X X_{ct} + \mathbf{\Gamma}_{ct}\boldsymbol{\gamma}_i + \varepsilon_{cit}, \quad i \in \{1, 79\}, \tag{16}$$

where α_c and α_t are country and year fixed effects, respectively, while $\mathbf{\Gamma}_{ct}$ is a vector of controls

[17] The coefficients in column (2) are identified through the implicit assumption that the errors of the two equations are jointly normal. In unreported specifications, we found similar results by estimating the Probit model with the lagged dependent variable as an additional regressor, which is excluded from the main equation. This variable has strong predictive power, consistent with it being a proxy for fixed export costs (Johnson, 2012). However, this variable may be correlated with unobserved determinants of trade, thereby violating the exclusion restriction.

Table 8: Sample Selection

	Inverse Hyperbolic Sine Transformation (1)	Heckman Correction (2)	Poisson Pseudo-Maximum Likelihood (3)
$S_{ct} \times AIint_i^r$	0.045**	0.053**	0.144***
	(0.019)	(0.022)	(0.037)
$I_{ct} \times AIint_i^r$	0.023***	0.028***	-0.016
	(0.008)	(0.009)	(0.021)
$X_{ct} \times AIint_i^r$	0.007***	0.003**	0.006**
	(0.001)	(0.001)	(0.003)
$G_{ct} \times AIint_i^r$	-0.003	-0.005	-0.003
	(0.004)	(0.005)	(0.011)
Controls	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes
Observations	51,184	49,535	51,184
Adj. R^2	0.696	0.663	0.793

Notes: In column (1), the dependent variable is the inverse hyperbolic sine transformation of exports to the US from country c in industry i and year t. In column (2), the dependent variable is the log of exports to the US from country c in industry i and year t. In column (3), the dependent variable is exports to the US from country c in industry i and year t, including observations with zero exports. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. The specification in column (3) is estimated using the Poisson pseudo-maximum likelihood estimator developed by Santos Silva and Tenreyro (2006). Controls include the interactions of skill and capital endowment of each country c in year t with the ranking of industries in terms of skill and capital intensity in 2020. Column (2) also includes the inverse Mills ratio, obtained using predicted values from a probit regression of a dummy equal to 1 for observations with non-zero exports on the set of regressors included in column (2). Standard errors, reported in parentheses, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

that includes the skill endowment, the capital endowment and the share of total new graduates in the population of each country c and year t, as in our preferred specification (column 7 of Table 5). The coefficients β_i^Z, identified from within-country variation over time, measure the relationship between a given source of CA, $Z = \{S, I, X\}$, and exports to the US in each industry i. With these coefficients in hand, we examine how they vary across industries with different levels of AI intensity. This approach allows us to study how AI intensity mediates the relation between a given country characteristic and exports, while remaining agnostic about the precise form of this relation.

The results are summarized in Table 9. The first three columns focus on the coefficients for the share of new STEM graduates in the population (β_i^S). The reported statistics are the mean and median values of these coefficients across all industries within a given tercile of the AI intensity distribution, along with the share of industries with $\beta_i^S > 0$ in each tercile. Interestingly, all these statistics increase steadily from the first to the third tercile of AI intensity, with an acceleration observed at the top of the distribution. The remaining columns of Table 9 show that similar patterns hold for the coefficients

Table 9: Non-Parametric Estimates

	New STEM Graduates per 100 Inhabitants (S_{ct})			Population Share with Internet Access (I_{ct})			Total Exports (X_{ct})		
	Mean β_i^S (1)	Median β_i^S (2)	% of Industries with $\beta_i^S > 0$ (3)	Mean β_i^I (4)	Median β_i^I (5)	% of Industries with $\beta_i^I > 0$ (6)	Mean β_i^X (7)	Median β_i^X (8)	% of Industries with $\beta_i^X > 0$ (9)
$AIint_i$ (Tercile 1)	0.295	-0.192	0.500	-0.220	-0.263	0.423	0.788	0.853	0.846
$AIint_i$ (Tercile 2)	0.531	0.888	0.630	-0.210	0.021	0.519	1.117	0.953	0.963
$AIint_i$ (Tercile 3)	1.311	1.360	0.769	0.062	0.141	0.654	0.950	0.968	0.846

Notes: The table shows the mean and median values of the coefficients β_i^S, β_i^I and β_i^X estimated from (16) across all industries within a given tercile of the distribution of AI intensity ($AIint_i$), as well as the share of industries with a positive estimate of a given coefficient in each tercile.

on the share of the population with Internet access (β_i^I) and on total exports (β_i^X). These findings confirm that the relationship between exports to the US and the three sources of CA is stronger in industries with higher AI intensity, irrespective of the functional form used. The fact that the strength of the relationship somewhat accelerates at the right tail of the AI intensity distribution implies that, if anything, the specification in (15) is likely to attenuate this relationship.

7 Threats to Identification

In this section, we discuss how confounding factors and reverse causality may threaten the identification of our parameters of interest.

7.1 Confounding Factors

As mentioned in Section 4, the industry-year fixed effects included in (15) capture all time-varying industry characteristics that uniformly affect exports across origin countries. The country-year fixed effects absorb instead all time-varying country characteristics that uniformly affect exports across industries. Our identification strategy relies on the assumption that, conditional on these fixed effects, the remaining variation in the regressors across country-industry-year triplets can be considered exogenous. In turn, this requires that, after controlling for the fixed effects, no time-varying country characteristic remains that correlates with the sources of CA and affects exports asymmetrically across industries with different AI intensity. In the following, we address the potential existence of such confounding factors and examine how they may affect the main results.

We start by considering various observable country characteristics. In column (1) of Table 10,

Table 10: Observable Confounders

	Other Determinants of CA	Additional Controls $\times$ $AIint_i^r$					Additional Controls $\times$ Industry FE				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
$S_{ct} \times AIint_i^r$	0.073***	0.051**	0.057***	0.045**	0.050**	0.050**	0.051**	0.062***	0.045**	0.050**	0.050**
	(0.026)	(0.022)	(0.022)	(0.022)	(0.022)	(0.023)	(0.022)	(0.021)	(0.022)	(0.022)	(0.023)
$I_{ct} \times AIint_i^r$	0.027**	0.026***	0.009	0.026***	0.026***	0.023**	0.025***	0.015	0.026***	0.027***	0.023**
	(0.011)	(0.010)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)
$X_{ct} \times AIint_i^r$	0.005***	0.004**	0.003**	0.004***	0.004***	0.005***	0.005***	0.003***	0.004***	0.004***	0.005***
	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.002)
$G_{ct} \times AIint_i^r$	-0.010	-0.005	-0.005	-0.003	-0.004	-0.002	-0.004	-0.005	-0.004	-0.005	-0.002
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	36,991	49,535	49,535	49,535	49,313	44,236	49,535	49,535	49,535	49,313	44,236
Adj. R^2	0.653	0.663	0.664	0.664	0.664	0.651	0.674	0.683	0.691	0.664	0.659

Notes: The dependent variable is the log of exports to the US from country c in industry i and year t. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. Controls include the interactions of skill and capital endowment of each country c in year t with the ranking of industries in terms of skill and capital intensity in 2020. In column (1), controls also include additional determinants of comparative advantage. These consist of: (i) the interactions of a proxy for the financial development of each country c in year t with two proxies for external finance dependence and asset tangibility of each industry i; and (ii) the interaction of a proxy for the institutional quality of each country c in year t with a proxy for contract intensity in each industry i. Columns (2)-(6) include, instead, the interactions between $AIint_i^r$ and the following characteristics of each country c in year t: real GDP, real per-capita GDP, growth in import penetration, the consumer price index, and the stock of US foreign direct investment over GDP. In columns (7)-(11), the same country characteristics are interacted with a full set of industry dummies rather than with $AIint_i^r$. Standard errors, reported in parentheses, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

we extend (15) by adding controls for other determinants of CA considered in the literature. Following, among others, Rajan and Zingales (1998), Manova (2013) and Bonfiglioli et al. (2019), we include interactions between a proxy for countries' financial development and proxies for industries' external financial dependence and asset tangibility. Following Nunn (2007) and Levchenko (2007), we also add the interaction between an index of countries' institutional quality and a proxy for the contract intensity of industries.[18] Although these variables are available only for manufacturing industries, controlling for them leaves the main results unchanged. This suggests that our evidence is not confounded by CA stemming from financial development or institutional quality.

In columns (2)–(6), we account for other country characteristics that may affect exports asymmetrically across industries. To this end, we extend the specification by adding interactions between $AIint_i^r$ and the following variables: real GDP, real per-capita GDP, growth in import penetration, the consumer price index and the stock of US foreign direct investment over GDP. The main evidence is preserved, suggesting that the results are not driven by the fact that larger, richer, more trade-exposed and higher-inflation economies—or countries with a greater presence of US multinational

[18]Financial development is proxied by the GDP share of private credit. External financial dependence and asset tangibility are industry i's rankings in terms of, respectively, the share of capital expenditure not financed by cash flow from operations and the share of net property, plant and equipment in total assets, both computed using firm-level data for the US. Institutional quality is proxied for using the rule of law, while contract intensity is industry i's ranking in terms of the indicator for the importance of relationship-specific investments constructed by Nunn (2007) for US industries.

Table 11: Underlying Trends and Contemporaneous Shocks

	Underlying Trends		Contemporaneous Shocks		
	(1)	(2)	(3)	(4)	(5)
$S_{ct} \times AIint_i^r$	0.047**	0.049**	0.054**	0.051**	0.051**
	(0.022)	(0.023)	(0.021)	(0.022)	(0.022)
$I_{ct} \times AIint_i^r$	0.027***	0.025***	0.025***	0.028***	0.027***
	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
$X_{ct} \times AIint_i^r$	0.003**	0.003**	0.004***	0.004***	0.004***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$G_{ct} \times AIint_i^r$	-0.003	-0.002	-0.003	-0.003	-0.004
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Controls	Yes	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes
Observations	49,535	48,386	49,535	49,535	49,535
Adj. R^2	0.668	0.659	0.685	0.665	0.663

Notes: The dependent variable is the log of exports to the US from country c in industry i and year t. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. Controls include the interactions of skill and capital endowment of each country c in year t with the ranking of industries in terms of skill and capital intensity in 2020. Columns (1)-(2) also control for the interaction between year dummies and the initial value of exports to (imports from) the US in each country c and industry i. Columns (3)-(4) also control for the interactions between year dummies and ten dummies for deciles of country-industry pairs, based on the growth in exports to (imports from) the US over the sample period. Column (5) also controls for interactions between dummies for years, four broad geographical areas (Europe and Central Asia, East Asia and Pacific, America and the Caribbean, Middle East and South Africa), and sectors (manufacturing and services). Standard errors, reported in parentheses, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

firms—might tend to export relatively more in AI-intensive industries. In columns (7)–(11), we interact the same country characteristics with a full set of industry dummies to more flexibly account for cross-industry differences in the effects of these variables. These interactions also absorb various determinants of country-industry-specific trade costs. The results are confirmed, even in these highly demanding specifications.

The large set of controls considered thus far may not completely eliminate concerns about confounding factors. On the one hand, it is possible that the results are driven by differential trends across country-industry pairs. To address this issue, in columns (1) and (2) of Table 11, we control for underlying trends based on pre-existing characteristics of each country-industry pair. To this end, in column (1), we interact the initial value of exports to the US from country c in industry i with a full set of year dummies. In column (2), we do the same using the initial value of US exports to country c in industry i. Reassuringly, the coefficients remain essentially unchanged.

On the other hand, the results may reflect unobserved time-varying shocks affecting specific country-industry pairs. We perform two complementary exercises to assess the role of these shocks. In columns (3)–(4) of Table 11, we divide country-industry pairs into ten bins corresponding to the deciles of the growth in exports to the US and in imports from the US, respectively, over the sample period. We then extend (15) by adding a full set of fixed effects for each decile-year pair. Conceivably, countries and industries with similar changes in trade with the US may have been affected by similar shocks during the sample period. These fixed effects are meant to absorb such shocks. Accordingly, identification is now achieved only from the remaining variation across countries and industries within the same bin and year. In column (5), we instead augment (15) with a full set of fixed effects for triplets of geographical areas, sectors and years.[19] These fixed effects account for all time-varying shocks affecting a specific sector within a region. Therefore, identification now relies solely on the remaining variation in the sources of CA across nearby countries, coupled with the residual variation in AI intensity across industries with similar technological content. The results remain unaffected in both cases.[20]

7.2 Reverse Causality

A second threat to our identification strategy is posed by reverse causality, i.e., the possibility that a change in exports to the US, occurring in a given country and industry for reasons unrelated to CA, affects the explanatory variables in a way that could determine the pattern of our coefficients.

We believe that, in practice, reverse causality is unlikely to be a key driver of the results. As shown before, our evidence holds across a wide range of proxies for CA. We view as unlikely that an export shock at the country-industry level could simultaneously alter all these country characteristics in a way that fully accounts for the results. At the same time, the AI intensity of industries is constructed using US data and kept constant over time; the stability of the ranking over the sample period suggests that it captures genuine technological differences across industries rather than transitory export shocks. Moreover, the results are robust across different proxies for AI intensity, based on alternative classifications of AI-related occupations. We view as improbable that an export shock could explain the cross-industry variation in all these proxies.

Nevertheless, to gauge the potential consequences of reverse causality on our findings, we perform two complementary exercises that exploit predetermined variation in the regressors. In columns

[19] Geographical areas are those listed in Table 3; sectors are manufacturing and services.

[20] In untabulated regressions, we have also augmented (15) with a full set of country-industry fixed effects to account for time-invariant heterogeneity across country-industry pairs. This specification relies solely on time variation for identification and is therefore less well-suited as a test of CA, which requires comparing different countries across different industries, as (15) does. The three sets of fixed effects jointly account for more than 95% of the variation in exports, leaving little variation to explain to the regressors. The results are qualitatively similar.

Table 12: Cross-Sectional Estimates

	OLS				2SLS			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$S_c \times AIint_i^r$	0.097***			0.061*	0.639***	0.700**	0.820***	1.565**
	(0.028)			(0.031)	(0.121)	(0.345)	(0.167)	(0.690)
$I_c \times AIint_i^r$		0.069***		0.024*		0.064**		0.118**
		(0.015)		(0.013)		(0.029)		(0.052)
$X_c \times AIint_i^r$			0.006***	0.005***		-0.001		-0.009
			(0.002)	(0.001)		(0.004)		(0.007)
$G_c \times AIint_i^r$	-0.023***			-0.008	-0.131***	-0.142*	-0.169***	-0.323**
	(0.008)			(0.008)	(0.027)	(0.073)	(0.036)	(0.144)
First-Stage Regression								
$T_c \times AIint_i^r$					0.009***	0.003***	0.007***	0.002***
					(0.000)	(0.001)	(0.001)	(0.000)
Kleibergen-Paap $F-$Statistic					651.84	26.65	169.31	20.02
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	3,499	5,070	5,070	3,499	2,263	2,263	2,263	2,263
Adj. R^2	0.689	0.688	0.689	0.692	0.700	0.697	0.675	0.516

Notes: The dependent variable is the log of average exports to the US from country c in industry i over the sample period. S_c is the number of new STEM graduates per 100 inhabitants aged 15-64 in the first sample year. I_c is the share of the population with Internet access in the first sample year. X_c is the log of total exports in the first sample year. G_c is the share of new graduates from all fields over the total population in the first sample year. $AIint_i^r$ is the ranking of industries by AI intensity in 2000. Controls include the interactions of skill and capital endowment of each country c in the first sample year with the ranking of industries in terms of skill and capital intensity in 2000. In columns (5)-(6) and (7)-(8), $S_c \times AIint_i^r$ is instrumented using the interaction between the total number of famous scientists (in thousands) born in each country c over the period 1780-1880 (T_c) and the value of $AIint_i^r$ in 2000 and 1980, respectively. Standard errors, reported in parenthesis, are corrected for clustering at the industry level. * Significant at 10%; ** significant at 5%; *** significant at 1%.

(1)–(4) of Table 12, we present cross-sectional results by regressing the log of average exports to the US in each country and industry over the sample period on the interactions of the initial values of S_{ct}, I_{ct} and X_{ct} with the first-year (2000) value of $AIint_i^r$. For consistency with (15), we control for country and industry fixed effects. The standard errors are corrected for clustering within industries. Because the explosion of AI technologies was largely unexpected at the beginning of the new millennium, the cross-sectional variation in the regressors in the initial year is likely to be predetermined to the evolution of exports in the two subsequent decades. Despite a significant loss of observations, the coefficients are similar to the baseline estimates.

In the second exercise, we leverage the fact that, for one of the three determinants of CA, it is possible to identify long-term historical differences across countries that predict its variation even in contemporary times. We exploit this variation for identification. This exercise will also provide additional interpretation of our results, as discussed below.

A large body of literature has highlighted that the spread of science within a country is strongly influenced by cultural values (see, e.g., Squicciarini, 2020; Lecce et al., 2021). To the extent that culture is persistent over time, current differences in the prevalence of new STEM graduates across countries are likely correlated with historical differences in the diffusion of science centuries ago. In turn, these historical differences are arguably uncorrelated with current shocks that could affect exports across countries and industries. Accordingly, we use long-term variation in the number of famous scientists across countries as an instrument for the share of new STEM graduates in the population.

We compute the total number of famous scientists born in each country from the outset of the Industrial Revolution (1780), which was fueled by the rapid growth in upper-tail human capital (Mokyr, 2005), until the late nineteenth century (1880, the last year with available data), using information on the number of famous individuals in scientific activities from de la Croix and Licandro (2015). We use this variable (T_c), interacted with the first-year (2000) value of $AIint_i^r$, as an instrument for the interaction between S_{ct} and $AIint_i^r$ in the previously estimated cross-sectional regression.

The results are reported in columns (5) and (6) of Table 12, first for a specification that only includes the share of new STEM graduates in the population as a source of CA and then for a specification that includes all three determinants of CA together. The first-stage coefficients are positive and very precisely estimated, pointing to the strong predictive power of the instrument.[21] At the same time, the 2SLS coefficients on the interaction between S_{ct} and $AIint_i^r$ remain positive and precisely estimated in both specifications. Hence, differences in the share of new STEM graduates in the population, driven by historical variation in the number of famous scientists across countries, are a significant determinant of countries' exports to the US, especially in AI-intensive industries.

In columns (7) and (8), we reconstruct the instrument by interacting the number of famous scientists (T_c) with the value of $AIint_i^r$ in 1980, i.e., twenty years prior to the beginning of the sample period, to further alleviate potential endogeneity concerns related to the cross-industry variation in AI intensity.[22] The main pattern of results is confirmed.

Comparing the OLS and 2SLS estimates in Table 12 shows that the 2SLS coefficients on the interaction between S_{ct} and $AIint_i^r$ are systematically larger than their OLS counterparts. As mentioned in Section 3.1.3, the share of new STEM graduates in the population captures not only the presence of advanced scientific skills—which are the skills that truly matter for CA in AI-intensive indus-

[21] The Kleibergen-Paap F-statistics for excluded instruments are well above 10, the value normally considered a rule-of-thumb threshold for instrument relevance.

[22] The value of $AIint_i^r$ in 1980 is computed using micro-level employment data from the 1980 US Census of Population. The cross-industry correlation between the 1980 and 2000 values of $AIint_i^r$ is very high (0.8), confirming that the ranking of industries by AI intensity captures technological differences that tend to persist over time.

tries—but also intermediate technical skills, which are less relevant. Using the number of famous scientists as an instrument corrects for this type of measurement error in S_{ct}, as it primarily isolates the cross-country variation in this variable that depends on differences at the top of the scientific skills distribution. Accordingly, the OLS coefficient likely provides a lower bound for the effect of these skills as a source of CA in AI-intensive industries.

8 The Role of Regulation

The question of whether and how governments can stimulate a country's export competitiveness is an old one. This question becomes particularly significant, however, for digitally-enabled trade, which may be less sensitive to traditional trade policy instruments. In this section, we provide evidence suggesting that various aspects related to the regulation of digitally-enabled trade may represent important sources of CA in AI-intensive industries. Unlike the determinants analyzed thus far, these regulatory factors can be directly and more rapidly influenced by governments to enhance national competitiveness. In this sense, the analysis in this section complements our earlier findings by offering insights into potential policy-relevant sources of CA in AI-intensive industries. Clearly, these policy interventions could be undertaken by governments in response to changes in international trade, so their variation may not be fully exogenous—a caveat to keep in mind when interpreting the results of this section.

We use data on the Digital Services Trade Restrictiveness Index (DSTRI) developed by the OECD and introduced in Section 3.2. The index ranges from 0 to 1, with higher values indicating a more regulated environment. We augment our main specification (15) by adding the interaction between DSTRI (henceforth labeled D_{ct}) and $AIint_i^r$. A negative coefficient on this interaction would imply that countries with a more liberal environment (a lower value of the index) export relatively more to the US in AI-intensive industries.

The results are reported in Table 13. In column (1), we consider a specification including only the interaction between D_{ct} and $AIint_i^r$; in column (2), we also include the interactions between $AIint_i^r$ and the three sources of CA related to scientific skills, digital infrastructure and economies of scale. The coefficient on the interaction between D_{ct} and $AIint_i^r$ is negative and precisely estimated in both cases, and its magnitude is largely unchanged across specifications. At the same time, the coefficients on the three interactions of $AIint_i^r$ with S_{ct}, I_{ct} and X_{ct} retain the same sign and approximately the same size as in the baseline specification; except for the coefficient on the STEM graduates-AI intensity interaction, whose p-value marginally exceeds 10% (0.114), these coefficients are also precisely estimated. These results suggest that the regulation of digitally-enabled trade operates beyond the

Table 13: The Role of Regulation

	Regulation Areas						
	All	All	Infrastructure & Connectivity	Electronic Transactions	E-Payment Systems	Intellectual Property Rights	Other Barriers
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$D_{ct} \times AIint_i^r$	-0.040**	-0.049**	-0.044*	0.133	-0.128	-0.492	-0.177*
	(0.016)	(0.023)	(0.024)	(0.112)	(0.165)	(0.318)	(0.098)
$S_{ct} \times AIint_i^r$		0.039	0.037	0.040	0.043*	0.025	0.048**
		(0.025)	(0.025)	(0.025)	(0.025)	(0.025)	(0.024)
$I_{ct} \times AIint_i^r$		0.038**	0.041**	0.048***	0.047***	0.049***	0.042***
		(0.016)	(0.016)	(0.015)	(0.015)	(0.015)	(0.015)
$X_{ct} \times AIint_i^r$		0.003**	0.003**	0.003**	0.004**	0.003**	0.004**
		(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$G_{ct} \times AIint_i^r$		-0.003	-0.002	-0.002	-0.003	-0.000	-0.005
		(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	23,076	16,366	16,366	16,366	16,366	16,366	16,366
Adj. R^2	0.631	0.671	0.671	0.670	0.670	0.671	0.670

Notes: The dependent variable is the log of exports to the US from country c in industry i and year t. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. D_{ct} is an index of the restrictiveness of regulations on digitally-enabled trade. Columns (1)-(2) use the composite index, while columns (3)-(7) use the sub-indices corresponding to regulation in five areas, as indicated in each column's heading. Controls include the interactions of skill and capital endowment of each country c in year t with the ranking of industries in terms of skill and capital intensity in 2020. Standard errors, reported in parentheses, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

role played by scientific skills, digital infrastructure and economies of scale.

Quantitatively, a less restrictive regulation, equivalent to the average difference in D_{ct} between Finland (25th percentile) and Singapore (75th percentile), is associated with 0.21 log points larger exports to the US in the industry at the 75th percentile of the AI intensity distribution relative to the industry at the 25th percentile. In addition to being statistically significant, therefore, the role of regulation is also quantitatively sizable and roughly comparable with that of the other sources of CA.

Along with the composite index, the OECD provides information on five sub-indices corresponding to the five areas of regulation encompassed by DSTRI: 1) infrastructure and connectivity; 2) electronic transactions; 3) e-payment systems; 4) intellectual property rights; and 5) other barriers to trade in digitally-enabled services. In the remaining columns of Table 13, we study the role of regulation in each of these areas. To this end, we replace the interaction between D_{ct} and $AIint_i^r$ with analogous interactions using the sub-indices of regulation for the individual areas.

The results indicate that not all dimensions of regulation are equally important. Exports in AI-intensive industries are particularly sensitive to regulations on infrastructure and connectivity, which include measures affecting cross-border data flows. Exports are also significantly influenced by other regulatory barriers, a residual category that encompasses performance requirements, limitations on downloading and streaming, and restrictions on online advertising, among other measures. The coefficients on the other interactions are estimated with imprecision. Hence, improving infrastructure and connectivity, as well as lifting restrictions on key aspects of business operations—such as performance requirements or online advertising—may serve as effective avenues for governments to enhance their countries' export competitiveness in AI-intensive industries.

9 Conclusions

This study has explored the determinants of CA in AI-intensive industries, leveraging a comprehensive dataset on US imports from 68 countries across 79 manufacturing and service industries over the last two decades. The findings reveal that countries with a higher availability of STEM graduates, broader Internet penetration and larger export volumes exhibit a CA in AI-intensive industries. Conversely, regulatory barriers to digital trade are negatively associated with export performance in these industries. These results are robust to various specifications and controls, and their economic magnitude is substantial.

We conclude by highlighting some limitations of our paper and proposing directions for future research. First, the adoption of AI technologies prior to 2020 does not encompass generative AI. Therefore, it remains an open question whether the most recent advancements in AI will alter the conclusions of our analysis in some way. Second, while this study leverages cross-industry and cross-country variation, substantial heterogeneity exists within industries as well. Given the significant differences in technology adoption rates across firms, exploring firm-level factors that drive AI diffusion would provide valuable insights. Third, the critical role of STEM workers in driving AI adoption suggests that advancements in AI may exacerbate inequality, even among skilled workers. Further work is needed to examine this potential disparity and identify possible remedies. Fourth, while the focus of this paper is on factors fueling AI-intensive industries, studying the effects of AI adoption on economic performance is an equally important question. Specifically, developing and estimating structural models could offer a deeper understanding of the impact of AI on global trade, productivity and labor markets.

Finally, more research is needed to inform the design of effective policies. As our results suggest, regulations can hinder AI adoption. Nevertheless, the growing integration of AI into business functions highlights the need for appropriate policies. A key challenge for the global economy is to enable countries to remain competitive in AI-intensive industries while avoiding an AI arms race and ensuring that developing nations are not left behind.

References

Acemoglu, D. and P. Restrepo (2020). Robots and jobs: Evidence from US labor markets. *Journal of Political Economy 128*(6), 2188–2244.

Autor, D. H. and D. Dorn (2013). The growth of low-skill service jobs and the polarization of the US labor market. *American Economic Review 103*(5), 1553–1597.

Autor, D. H., D. Dorn, and G. H. Hanson (2013). The China syndrome: Local labor market effects of import competition in the United States. *American Economic Review 103*(6), 2121–68.

Bailey, M., A. Gupta, S. Hillenbrand, T. Kuchler, R. J. Richmond, and J. Stroebel (2021). International trade and social connectedness. *Journal of International Economics 129*, 103418.

Balassa, B. (1965). Trade liberalisation and "revealed" comparative advantage. *The Manchester School 33*(2), 99–123.

Blum, B. S. and A. Goldfarb (2006). Does the internet defy the law of gravity? *Journal of International Economics 70*(2), 384–405.

Bonfiglioli, A., R. Crinò, and G. Gancia (2019). Trade, finance, and endogenous firm heterogeneity. *Journal of the European Economic Association 17*(1), 79–130.

Bonfiglioli, A., G. Gancia, I. Papadakis, and R. Crinò (2025). Artificial Intelligence and jobs: Evidence from US commuting zones. *Economic Policy*, forthcoming.

Brynjolfsson, E., X. Hui, and M. Liu (2019). Does machine translation affect international trade? Evidence from a large digital platform. *Management Science 65*(12), 5449–5460.

Carballo, J., M. Rodriguez Chatruc, C. Salas Santa, and C. Volpe Martincus (2022). Online business platforms and international trade. *Journal of International Economics 137*, 103599.

Chen, M. X. and M. Wu (2021). The value of reputation in trade: Evidence from Alibaba. *Review of Economics and Statistics 103*(5), 857–873.

Chor, D. (2010). Unpacking sources of comparative advantage: A quantitative approach. *Journal of International Economics 82*(2), 152–167.

Costinot, A., D. Donaldson, and I. Komunjer (2012). What goods do countries trade? A quantitative exploration of Ricardo's ideas. *Review of Economic Studies 79*(2), 581–608.

de la Croix, D. and O. Licandro (2015). The longevity of famous people from Hammurabi to Einstein. *Journal of Economic Growth 20*, 263–303.

Demir, B., B. Javorcik, and P. Panigrahi (2025). Breaking invisible barriers: Does fast internet improve access to input markets? *CEPR Discussion Paper 19827*.

Eaton, J. and S. Kortum (2002). Technology, geography, and trade. *Econometrica 70*(5), 1741–1779.

Feenstra, R. C., J. Romalis, and P. K. Schott (2002). US imports, exports, and tariff data, 1989-2001. *NBER Working Paper 9387*.

Ferencz, J. (2019). The OECD Digital Services Trade Restrictiveness Index. *OECD Trade Policy Papers 221*.

Ferencz, J., J. L. González, and I. O. García (2022). Artificial intelligence and international trade. *OECD Publishing 260*.

Freund, C. L. and D. Weinhold (2004). The effect of the internet on international trade. *Journal of International Economics 62*(1), 171–189.

Goldfarb, A. and D. Trefler (2019). Artificial intelligence and international trade. In A. Agrawal, J. Gans, and A. Goldfarb (Eds.), *The Economics of Artificial Intelligence: An Agenda*.

Hanson, G. H. and M. J. Slaughter (2018). High-skilled immigration and the rise of STEM occupations in US employment. In *Education, Skills, and Technological Change: Implications for future US GDP Growth*, pp. 465–494. NBER Books: National Bureau of Economic Research.

Heckman, J. (1979). Sample selection bias as a specification error. *Econometrica 47*(1), 153–161.

Johnson, R. C. (2012). Trade and prices with heterogeneous firms. *Journal of International Economics 86*(1), 43–56.

Lecce, G., L. Ogliari, and M. P. Squicciarini (2021). Birth and migration of scientists: Does religiosity matter? Evidence from 19th-century France. *Journal of Economic Behavior & Organization 187*, 274–289.

Lendle, A., M. Olarreaga, S. Schropp, and P.-L. Vézina (2016). There goes gravity: eBay and the death of distance. *Economic Journal 126*(591), 406–441.

Levchenko, A. (2007). Institutional quality and international trade. *Review of Economic Studies 74*(3), 791–819.

Manova, K. (2013). Credit constraints, heterogeneous firms, and international trade. *Review of Economic Studies 80*(2), 711–744.

Mokyr, J. (2005). Long-term economic growth and the history of technology. In *Handbook of Economic Growth*, Volume 1, pp. 1113–1180. Elsevier.

Nunn, N. (2007). Relationship-specificity, incomplete contracts, and the pattern of trade. *The Quarterly Journal of Economics 122*(2), 569–600.

Rajan, R. and L. Zingales (1998). Financial dependence and growth. *American Economic Review 88*(3), 559–86.

Romalis, J. (2004). Factor proportions and the structure of commodity trade. *American Economic Review 94*(1), 67–97.

Ruggles, S., S. Flood, M. Sobek, D. Backman, A. Chen, G. Cooper, S. Richards, R. Rogers, and M. Schouweiler (2023). *IPUMS USA: Version 14.0*. Minneapolis, MN: IPUMS, 2023.

Santos Silva, J. and S. Tenreyro (2006). The log of gravity. *Review of Economics and Statistics 88*(4), 641–658.

Squicciarini, M. P. (2020). Devotion and development: Religiosity, education, and economic progress in nineteenth-century France. *American Economic Review 110*(11), 3454–3491.

Sun, R. and D. Trefler (2022). AI, trade, and creative destruction: A first look. In L. Y. Ing and G. M. Grossman (Eds.), *Robots and AI: a New Economic Era*. Oxon and New York: Routledge.

Sun, R. and D. Trefler (2023). The impact of AI and cross-border data regulation on international trade in digital services: A large language model. *NBER Working Paper 31925*.

Webb, M. (2020). The impact of artificial intelligence on the labor market. *Stanford University*, Mimeo.

A Data and Variables

In this appendix, we provide detailed information about data sources and variable definitions. Our sample covers the period from 1999 to 2019 and includes 79 industries, of which 66 belong to the manufacturing sector and 13 to the service sector. The number of countries ranges from 45 to 68, depending on the source of CA considered in the analysis, as shown in Table A.1.

A.1 Exports

Our dependent variable, M_{cit}, is exports to the US from country c in industry i and year t. The data come from two sources. For the 66 manufacturing industries, we use bilateral data on US imports from Feenstra et al. (2002); we aggregate the product-level data, originally classified according to the 10-digit level of the Harmonized System (HS) classification, at the 4-digit level of the North American Industry Classification System (NAICS). For the 13 service industries, we use bilateral data on US imports from the Bureau of Economic Analysis (BEA).

A.2 AI Intensity of Industries

To measure the AI intensity of each industry, we first identify a set of 19 AI-related occupations, as explained in Section 3.1.2, using information on the software requirements most frequently included in all current employer job postings in the US from the "Hot Technologies" section of the O*NET database. The list of software used to identify the AI-related occupations is provided in Table A.2. Next, we combine the list of AI-related occupations with micro-level employment data from the 2020 American Community Survey (ACS), sourced from Ruggles et al. (2023). The ACS represents a 1% random sample of the US population. To increase sample size, we follow Autor et al. (2013) and Acemoglu and Restrepo (2020) by using pooled 5-year ACS data for 2017-2021. We define the AI intensity of each industry i, $AIint_i$, as the ratio of employment in AI-related occupations to employment in non-AI-related occupations, both computed using sample weights. We focus on workers aged 16-65 who were employed the year prior to the interview and who do not reside in institutional group quarters, following Autor and Dorn (2013). We use the same approach to measure AI intensity in 2000 and 1980, using micro-level employment data from the US Census of Population for the corresponding years. In both cases, we use a 5% random sample. To track AI-related occupations back in time, across the revisions of the Standard Occupational Classification (SOC) system that occurred between 1980 and 2020, we utilize correspondence tables from the US Bureau of Labor Statistics (BLS).

Table A.1: Countries Included in the Sample

Region	Countries
Europe and Central Asia	Austria, Belgium, Bulgaria, Croatia, Cyprus*, Czechia, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta*, Netherlands, Norway, Poland, Portugal, Romania, Russia, Slovakia, Slovenia, Spain, Sweden, Switzerland, Turkey, United Kingdom
East Asia and Pacific	Australia, Brunei*, China*, Hong Kong*, India*, Indonesia, Japan, Malaysia*, New Zealand, Philippines*, Singapore*, South Korea, Thailand*, Vietnam*
America and Caribbean	Argentina, Bermuda*, Brazil, Canada, Chile, Colombia, Costa Rica, Dominican Republic*, El Salvador*, Guatemala*, Honduras*, Mexico, Nicaragua*, Panama*, Peru*
Middle East and Africa	Bahrain*, Israel, Jordan*, Morocco*, Oman*, Saudi Arabia, South Africa

Notes: Geographical areas are defined according to the World Bank definition. *Countries with no data on new STEM graduates.

A.3 Sources of Comparative Advantage in AI-Intensive Industries

We measure the availability of scientific skills in a given country c and year t using information on new bachelor's, master's, and Ph.D. graduates in STEM fields (science, technology, engineering and mathematics) from the OECD Education Statistics "Graduates by Fields" database. S_{ct} is calculated as the number of new graduates in STEM fields per 100 inhabitants aged 15-64, using data on the total country population from the Penn World Tables and on the share of the population aged 15-64 from the World Bank World Development Indicators (WDI). We use the same data and approach to compute: (i) the share of new total graduates in the population aged 15-64, G_{ct}; and (ii) the number of new STEM graduates by field (Natural Sciences, Mathematics and Statistics, Engineering and ICT) per 100 inhabitants aged 15-64.

Digital infrastructure is proxied by the share of the population with Internet access in each country c and year t (I_{ct}), sourced from the WDI. X_{ct} represents the log of total exports from each country c in year t. Total exports are calculated by multiplying the GDP share of exports from the WDI by each country's GDP, computed using data on per-capita GDP from the WDI and on total country population from the Penn World Tables.

A.4 Other Determinants of Comparative Advantage

For each country c and year t, we proxy skill endowment using the log index of human capital (average years of schooling per person) and capital endowment using the log of real capital stock per person engaged. Both measures are sourced from the Penn World Tables. The skill intensity of industry i is

Table A.2: Software Used to Identify the AI-Related Occupations

Amazon Redshift	GitHub	Oracle PL/SQL
Amazon Simple Storage Service S3	Go	PHP
Amazon Simple Storage Service S4	JavaScript	Perl
Amazon Web Services AWS CloudFormation	JavaScript Object Notation JSON	PostgreSQL
Amazon Web Services AWS Software	Jenkins CI	Python
Ansible Software	Kubernetes	Ruby
Apache Hadoop	Microsoft NET Framework	Seala
Apache Hive	Microsoft Azure Software	Selenium
Apache Kafka	Microsoft PowerShell	ServiceNow
Apache Spark	Microsoft SQL Server	Splunk Enterprise
Atlassian Confluence	Microsoft SQL Server Reporting Services SSRS	Spring Boot
Atlassian JIRA	MongoDB	Spring Framework
Bash	NoSQL	Structured Query Language SQL
C	Node.js	Transact-SQL
C#	Objective C	TypeScript
C++	Oracle Database	UNIX
Docker	Oracle Java	Vue.js
Git	Oracle Java 2 Platform Enterprise Edition J2EE	jQuery

defined as the log ratio of the number of employees with at least a bachelor's degree to the number of employees with less than a bachelor's degree, computed using 2020 ACS data. Capital intensity is defined as the log ratio of capital compensation to labor compensation, with data sourced from the NBER Manufacturing Industry Productivity database for the 66 manufacturing industries and from the Production Account Tables (PAT) of the BEA for the 13 service industries.

The financial development of each country c in year t is defined as the GDP share of private credit, sourced from the WDI. Industry i's external finance dependence and asset tangibility are defined, respectively, as the share of capital expenditure not financed by cash flow from operations and as the share of net property, plant and equipment in total assets. Both measures are computed using firm-level data for the US from Compustat and refer to the median values of each industry for the year 2010. The institutional quality of each country c is proxied by the average rule of law, sourced from the World Bank Worldwide Governance Indicators database. Industry i's contract intensity is measured by the importance of relationship-specific investments, as constructed by Nunn (2007) using data for the US in the year 1997.

A.5 Controls for Confounding Factors

The time-varying country characteristics used in Table 10 are: (i) real GDP at constant 2017 prices in US$ million from the Penn World Tables; (ii) real per-capita GDP at constant 2015 prices in US$ from the WDI; (iii) growth in import penetration over the sample period, calculated using the GDP share of imports of goods and services from the WDI; (iv) the consumer price index from the WDI; and (v) the stock of US foreign direct investment as a percentage of real GDP, sourced from the BEA - Balance of Payments and Direct Investment Position database. In Table 11, the log of US exports to the 68 countries in each industry i is calculated using bilateral data on US exports from Feenstra et al. (2002) for the 66 manufacturing industries and from the BEA for the 13 service industries.

A.6 Alternative Measures of AI Intensity and Scientific Skills

The alternative measures of AI intensity used in columns (2)-(4) of Table 7 are constructed as follows. First, we use the log ratio of employment in all STEM occupations to employment in non-STEM occupations, based on the definition of STEM occupations provided by Hanson and Slaughter (2018). Second, we use the same measure but exclude life and medical scientists as well as other STEM occupations, such as surveyors, cartographers and mapping scientists (codes 191010, 191020, 191040, 1910XX and 173031 of the OCCSOC variable in the ACS). Third, we use the log ratio of employment in occupations in the top quartile of the AI exposure index proposed by Webb (2020) to employment in all other occupations. All three alternative measures are computed using micro-level employment data from the 2020 ACS. In column (9) of Table 7, we replace the endowment of scientific skills, measured by the number of new STEM graduates per 100 inhabitants aged 15-64, with the ICT share of capital stock. The latter is defined as the share of communication and computer equipment in total assets and is sourced from the EU KLEMS database.

A.7 Digital Services Trade Restrictiveness Index

The Digital Services Trade Restrictiveness Index (DSTRI) is developed by the OECD to measure the restrictiveness of regulations affecting digital services trade and data flows across countries. The DSTRI is a composite index that takes values between 0 and 1, where 0 indicates an open regulatory environment for digitally-enabled trade and 1 indicates a completely closed regime. As reported in Ferencz (2019), the DSTRI encompasses regulations in the following five areas.

- *Infrastructure and connectivity*. This area covers measures related to communication infrastructures essential to engaging in digital trade. It maps the extent to which best practice regulations on interconnections among network operators are applied to ensure seamless communication. It also captures measures limiting or blocking the use of communications services, including Virtual Private Networks or leased lines. Lastly, this area covers policies that affect connectivity such as measures on cross-border data flows and data localization.

- *Electronic transactions*. This area covers issues such as discriminatory conditions for issuing licenses for e-commerce activities, the possibility for online tax registration and declaration for non-resident firms, deviation from internationally accepted rules on electronic contracts, measures inhibiting the use of electronic authentication (such as electronic signature), and the lack of effective dispute settlement mechanisms.
- *E-payment systems*. This area captures measures that affect payments made through electronic means. It includes measures related to access to certain payment methods and assesses whether domestic security standards for payment transactions are adopted in line with international standards. Lastly, it also covers restrictions related to Internet banking not covered in other areas.
- *Intellectual property rights*. This area covers domestic policies related to copyrights and trademarks that do not afford foreigners equal treatment with regard to IP protection. It also maps the existence of appropriate enforcement mechanisms to address infringements related to copyrights and trademarks, including those occurring online.
- *Other barriers affecting trade in digitally enabled services*. This area covers various other barriers to digital trade, including performance requirements affecting cross-border digital trade (e.g., mandatory use of local software and encryption or mandatory technology transfers); limitations on downloading and streaming; restrictions on online advertising; commercial or local presence requirements; and lack of effective redress mechanisms against anti-competitive practices online, among others.

A.8 Instrument for New STEM Graduates

The instrument T_c used in Table 12 is the number of famous individuals in scientific activities who were born in each country c from 1780 until 1880. These data are sourced from de la Croix and Licandro (2015).

B Additional Results

Table B.1: AI Intensity by 4-Digit NAICS Industry

Industry	$AIint_i$	Industry	$AIint_i$
Information and Data Processing Services	0.2633	Aluminium Production and Processing	0.0158
Computer Equipment Manufacturing	0.2190	Other Food Manufacturing	0.0157
Communications, Audio and Video Equipments	0.1415	Beverage	0.0156
Other Business Services	0.1363	Structural Metals, and Tank and Shipping Container and Boiler Manufacturing	0.0154
Telecommunication Services	0.1305	Paper Product Manufacturing	0.0149
Electronic Components and Products	0.1015	Other Fabricated Metal Product Manufacturing	0.0142
Navigational Electronical and Control Instruments Manufacturing	0.1000	Rubber Products	0.0139
Financial Services	0.0953	Textile and Fabric Finishing	0.0136
Audiovisual Services	0.0915	Furniture, Fixtures and Related Products	0.0134
Aerospace Products Manufacturing	0.0871	Plastic Products	0.0133
Insurance Services	0.0765	Fabric Mills	0.0133
Pharmaceuticals and Medicines	0.0563	Cut and Sew Manufacturing and Apparel Accessories	0.0132
Medical Equipment and Supplies Manufacturing	0.0523	Clay Product and Refractory Manufacturing	0.0131
Commercial and Service Industry Machinery Manufacturing	0.0479	Carpet and Rug Mills and Manufacturing	0.0129
Engine, Turbine, Power Transmission Equipment Manufacturing	0.0429	Health Services	0.0125
Railroad Rolling Stock Manufacturing	0.0408	Fruit and Vegetable Preserving and Speciality Food	0.0121
Agriculture, Construction, Mining Machinery Manufacturing	0.0350	Iron, Steel Mills and Steel Products Manufacturing	0.0120
Water Transportation	0.0344	Sugar and Confectionery Products	0.0119
Household Appliance Manufacturing	0.0333	Artistic-Related Services	0.0114
Electrical Machinery, Equipment, and Other Electrical Components	0.0330	Metal Forgins and Stamping	0.0111
Petroleum Refining	0.0321	Resin, Synthetic Rubber, Fibers	0.0105
Other General Purpose Machinery Manufacturing, Nec, or Not Specified	0.0320	Machine Shops	0.0105
Motor Vehicles and Motor Vehicle Equipment	0.0281	Other Modes of Transport	0.0104
Paint, Coating, Adhesives	0.0279	Veneer, Plywood, Engineered Wood Product	0.0100
Air Transportation	0.0266	Leather and Hide Tanning and Finishing and Other Allied Products	0.0097
Printing and Related Support Activities	0.0252	Textile Product Mills Except Carpet and Rugs	0.0094
Other Miscellaneous Manufacturing	0.0228	Glass and Glass Products	0.0091
Footwear Manufacturing	0.0223	Nonferrous Metals	0.0090
Industrial and Miscellaneous Chemicals	0.0222	Foundries	0.0089
Cutlery and Hand Tool Manufacturing	0.0220	Dairy Products	0.0089
Ship and Boat Manufacturing	0.0220	Seafood and Other Miscellaneous Foods, Nec	0.0088
Agricultural Chemicals	0.0219	Sawmills and Wood Preservation	0.0087
Knitting Mills	0.0219	Miscellaneous Nonmetallic Mineral Products	0.0086
Soap, Cleaning Compound, Cosmetics	0.0216	Other Woods Products	0.0085
Pulp, Paper and Paperboard Mills	0.0213	Cement, Concrete, Lime, and Gypsum Products	0.0074
Educational Services	0.0211	Animal Slaughtering and Processing	0.0070
Other Transportation Equipment Manufacturing	0.0195	Bakeries	0.0031
Animal Food, Grain and Oilseed Milling	0.0189	Construction	0.0030
Tobacco	0.0177	Fiber, Yarn, and Thread Mills	0.0028
Metalworking Machinery Manufacturing	0.0162		

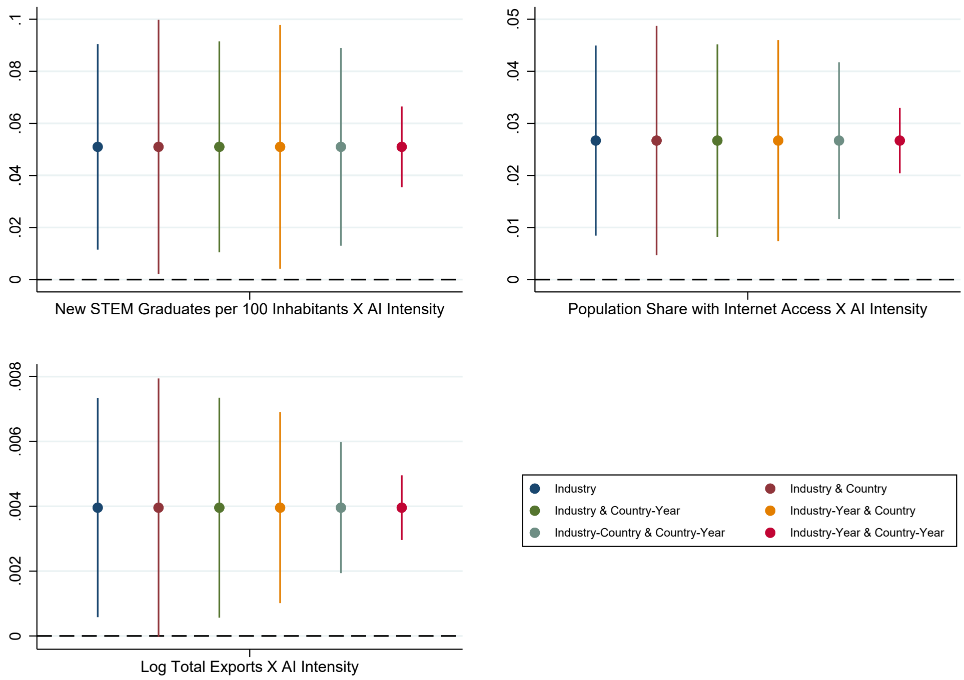
Notes: $AIint_i$ is the ratio of employment in AI-related occupations to employment in non-AI-related occupations in each industry in 2020. AI-related occupations are listed in Table 1. Industries are classified according to the NAICS 4-digit classification.

Table B.2: Baseline Estimates with Controls

	(1)	(2)	(3)	(4)	(5)	(6)
$S_{ct} \times AIint_i^r$	0.047***		0.072***			0.051**
	(0.014)		(0.022)			(0.022)
$I_{ct} \times AIint_i^r$				0.058***		0.027***
				(0.007)		(0.009)
$X_{ct} \times AIint_i^r$					0.007***	0.004***
					(0.001)	(0.001)
$G_{ct} \times AIint_i^r$		0.006	-0.008			-0.004
		(0.003)	(0.005)			(0.005)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Country-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry-Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	49,812	49,887	49,742	92,120	92,327	49,535
Adj. R^2	0.660	0.660	0.661	0.627	0.626	0.663

Notes: The dependent variable is the log of exports to the US from country c in industry i and year t. S_{ct} is the number of new STEM graduates per 100 inhabitants aged 15-64. I_{ct} is the share of the population with Internet access. X_{ct} is the log of total exports. G_{ct} is the share of new graduates from all fields over the total population. $AIint_i^r$ is the ranking of industries by AI intensity in 2020. Controls include the interactions of skill and capital endowment of each country c in year t with the ranking of industries in terms of skill and capital intensity in 2020. Standard errors, reported in parenthesis, are corrected for two-way clustering by country-industry and industry-year. * Significant at 10%; ** significant at 5%; *** significant at 1%.

Figure B.1: Alternative Clustering



Notes: Dots correspond to the baseline coefficients (Table 5, column 7). All confidence intervals are at the 90% level and correspond to standard errors corrected using the clustering schemes indicated in the legend.