A.I. in the Office and the Factory: Evidence from Administrative Software Registry Data*

Gustavo de Souza†

June 2026

Abstract

What is the effect of artificial intelligence (AI) on employment and wages? Using administrative data on a comprehensive set of AI applications developed in Brazil, I show that AI is widely used not only in administrative tasks but also in the production setting, where it primarily supports process optimization and quality control. Exploiting exogenous variation in AI development costs, I find that AI affects administrative and production workers differently. In office occupations, AI reduces employment, especially among young workers performing data-processing tasks. In production occupations, it increases employment of young and low-skilled workers by helping them perform decision-making tasks that previously required expertise. On net, AI increases employment in Brazil because gains in production outweigh losses in office work.

Key Words: Artificial Intelligence, Automation, Deskilling, Software, Inequality

JEL: J23, J24, F63

* This paper benefited from valuable comments from Ajay Agrawal (discussant), Francesco Amodio, Michael Andrews, David Autor, Wyatt Brooks, Bharat Chandar, Raj Chetty, Oren Danieli, John Friedman, Tarek Hassan, Nathaniel Hendren, Seyed M. Hosseini, Anders Humlum, Xavier Jaravel, Joe Kaboski, Aakash Kalyani, David Lagakos, Kristina McElheran, Markus Poschke, Diego Restuccia, Matthew Schwartzman, Anna Salomons (discussant), Yongseok Shin, and Kjetil Storesletten. I also thank seminar and conference participants at the Opportunity Insights Conference on Economic Mobility, Boston University, McGill University, the Federal Reserve Bank of San Francisco, the Federal Reserve Bank of Chicago, the IMF, LACEA, SBE, MEA, the Artificial Intelligence and the Macroeconomy Conference, the NBER AI and Economic Measurement Conference, HKU, HKUST, CUHK, Tilburg University, and the STEG Annual Conference. Jack Mannion and Vitor Santos provided superb research assistance. I thank Haishi Li and Vitor Santos for providing the survey data.

† Federal Reserve Bank of Chicago. Contact: gustavo@microtomacro.net. This paper represents my opinions and not those of the Federal Reserve Bank of Chicago or the Federal Reserve System.

"I like to think (and
the sooner the better!)
of a cybernetic meadow
where mammals and computers
live together in mutually
programming harmony."

—Richard Brautigan,
"All Watched Over by Machines of Loving Grace"

1 Introduction

Artificial intelligence (AI) refers to computer systems that learn from data to recognize patterns, make predictions, or generate information—tasks that once required human labor. Advances in hardware and the widespread collection of digital data have made AI a quiet, everyday technology. It runs social media feeds, supports banking transfers, guides self-driving cars, automates industrial processes, and powers chatbots, among many other uses (Marr and Ward 2019). This breadth of applications has drawn public attention to AI's potential effects on the labor market. One view holds that AI automates tasks and displaces workers, especially in routine cognitive jobs (Acemoglu and Restrepo 2019, Webb 2020, Aum and Shin 2025). A contrasting perspective argues that AI can augment labor by simplifying tasks and improving decision-making, ultimately creating jobs (Autor 2024, Agrawal et al. 2019b, Bresnahan 2021). Despite growing interest, there is still no consensus on whether AI is, on net, decreasing or increasing employment.

A possible reason for these contrasting views on the effects of AI is that much of the existing literature studies AI in specific settings, reaching different conclusions. Some focus on generative AI tools used by professionals (Humlum and Vestergaard 2025b, Brynjolfsson et al. 2025), while others focus on the effects of predictive AI in firms that develop their own technologies in-house (McElheran et al. 2023, Brynjolfsson and McElheran 2019, Hampole et al. 2025, Babina et al. 2024, Adams et al. 2024). Yet, evidence suggests that the scope of AI adoption is broader. AI is increasingly embedded in third-party software, reaching firms through off-the-shelf tools rather than internal development (Aoun et al. 2022, Xu et al. 2018, Li et al. 2024). Engineers highlight a wider set of applications, including factory settings, where AI appears to raise machine productivity and operational efficiency (McElheran et al. 2025, Alam et al. 2024, Marr and Ward 2019). This diversity of applications could lead to different conclusions because the effects of AI depend on how the technology is applied

(Aghion et al. 2025, Ide and Talamàs 2025). Therefore, to understand AI's overall labor-market impact, we need data that captures a representative range of AI applications.

I contribute to this debate by studying how AI affects the labor market using an administrative registry of commercial AI software developed in Brazil. Each registration records technical descriptions and intended uses reported by the firm that developed the software and verified by an external examiner during the registration process. The data cover a comprehensive set of AI applications, from early predictive systems to generative AI tools registered by 2024. This breadth allows me to study AI across a wide range of real-world uses.

Using exogenous variation in AI development costs as an instrument, I find that AI affects administrative and production workers in different ways. AI reduces employment in administrative occupations, displacing mostly young workers performing tasks related to data processing, such as data entry, form completion, or recordkeeping. In contrast, among production workers, AI increases employment of young, inexperienced, and low-education workers. More generally, AI expands employment in occupations that use operational information to guide decisions and actions, such as production work, while replacing occupations centered on gathering and processing information, as theorized by Agrawal et al. (2019b). In aggregate, due to the composition of the Brazilian economy, employment gains among production workers dominate, resulting in higher employment and lower inequality.

These contrasting effects help reconcile competing views on the labor-market impact of AI. In production, the results align with the idea that AI increases machine productivity, simplifies operation, and raises demand for low-skilled labor (Autor 2024, Bresnahan 2021, Agrawal et al. 2019b). In office settings, they are consistent with AI automating tasks and reducing employment (Acemoglu and Restrepo 2018, Webb 2020, Aum and Shin 2025, Hui et al. 2024). Taken together, the findings suggest that the impact of AI depends on how and where it is used, explaining why prior studies, focused on specific applications or firms, reached different conclusions.

To see these effects in action, consider Techplus, a Brazilian firm that develops AI systems for industrial operations. Its software, which is part of Brazil's software registry, analyzes real-time sensor data from factory equipment to predict failures, schedule maintenance, and assist machine operation. In doing so, the software provides real-time information and expert knowledge to supervisors and machine operators. Such systems can reduce unplanned downtime by up to 50% (Agoro 2025, Benhanifia et al. 2025), which raises the value of keeping

machines running and increases demand for workers who operate them (Schmitz 2005). At the same time, the software automates routine administrative tasks such as maintenance planning and inventory control, which is likely to reduce the demand for workers performing these information-processing tasks. This case reflects the broader pattern I document in the data: AI increases employment in low-skilled production roles by making expertise available, while reducing demand for routine office work through automation.

To study the effect of AI on the labor market, I assemble a dataset on AI software developed in Brazil. In 1987, Brazil introduced copyright protection for the source code of computer programs registered with the National Institute of Industrial Property (NIIP). Firms register software to establish legal proof of ownership, protect source code against copying, and support claims in legal disputes. As a result, registration became standard industry practice: approximately 92% of software development firms in 2016 had at least one program registered.[1] To register a program, firms report its title, intended use, and technological features, including whether it uses AI methods such as neural networks or deep learning. An NIIP examiner then evaluates this information and, if it is confirmed, the firm receives a certificate of registration that can be used in legal proceedings. I collect data on all registered programs that use AI technology. Because NIIP records contain verified technical and application details on deployed software, they offer a clearer measure of AI development than patents or job advertisements, the standard proxies in the existing literature.[2]

The NIIP registry shows two relevant facts about AI development. First, AI is not confined to administrative tasks. While 48 percent of AI software targets office or clerical functions, 44 percent is designed for core production activities in manufacturing, agriculture, construction, and other sectors. Second, production-oriented AI is concentrated in a few practical applications: 40 percent are in software to manage production, 21 percent handle automated quality inspection, and 15 percent are industrial robot interface software. The intensive use of AI in production challenges the idea that AI affects only white-collar work and points instead to substantial labor-market implications for machine operators and other

[1] To assess the coverage of the registry, I also conducted a survey of industry specialists. Ninety-six percent of respondents are aware of the NIIP registration process or have taken part in it directly. Seventy-one percent report that most or all firms register at least one version of their software.

[2] The main limitation of this dataset is that it does not cover software developed abroad. However, due to regulatory barriers and a strong home bias in computer services (Alaveras and Martens 2015, Miroudot et al. 2010, Harms and Shuvalova 2020, Ezell et al. 2013, Benz and Jaax 2022), foreign software plays a minor role in the Brazilian market. In 2015, imported computer services accounted for only 9.3% of domestic consumption, less than half the OECD average of 19%, and fewer than 0.02% of IT firms were foreign-owned. As a result, the majority of software used in Brazil is developed locally and thus captured by the NIIP registry.

shop-floor workers.[3]

To estimate how AI affects different occupations, I calculate an occupation-level AI exposure over time. The AI exposure is constructed using text similarity between the tasks performed by each occupation and the descriptions of AI software in the NIIP database, following standard approaches in the literature (Webb 2020, de Souza and Li 2023, Kogan et al. 2023, Hampole et al. 2025). The intuition is simple: occupations whose tasks resemble those performed by AI software are more exposed to AI development, which can either complement or substitute their work. By conducting the analysis at the occupation level, the AI exposure captures workers' exposure to both in-house and vendor-supplied AI and the resulting reallocation of workers across firms and sectors.[4]

To estimate the causal effect of AI, I use an instrumental variable strategy that exploits exogenous variation in the cost of developing AI software. The instrument relies on the fact that software development costs depend on the availability of resources for a programming language, such as libraries, tutorials, and community support, allowing developers to reuse scripts rather than build them from scratch (Delorey et al. 2007, Krein et al. 2010, Tambad et al. 2020, Vijayaraghavan et al. 2022). The identifying assumption is a parallel-trends condition: absent these language-specific cost shocks, Brazilian occupations more and less exposed to the instrument would have followed similar labor-market trends. I support this assumption by showing that the instrument is not systematically related to development in other Brazilian technologies, to the price of imported equipment, to sectoral trade shocks, or lagged labor-market outcomes.

I find that AI increases employment while drawing in workers with lower observable and unobservable skills. A one-standard-deviation increase in AI exposure raises occupational employment by about 3%, with stronger gains among younger, less educated, less experienced, and lower-ability workers. After conditioning on worker characteristics, AI increases

[3]Although the use of AI in production settings has received little attention from economists, it is widely studied in the engineering and operations literature. For example, Chen et al. (2006), Capgemini (2020), and Shojaeinasab et al. (2022) document the integration of AI into Manufacturing Execution Systems (MESs). AI is used in quality control for detecting anomalies and improving inspection accuracy, as discussed by Ghelani et al. (2024), Tariq et al. (2024), Elsafty et al. (2024), Shi et al. (2023), Yang et al. (2023), and Patel and Ghelani (2024), among many others. Additionally, AI is used to optimize energy usage in manufacturing (Evans and Gao 2016, Liu et al. 2022). Finally, Marr and Ward (2019) and Borboni et al. (2023) discuss the use of AI in the operation of industrial robots.

[4]I conduct the analysis at the occupation level because 81% of software developed in Brazil is produced by specialized IT firms and licensed to many clients; a firm-level analysis would therefore miss much AI adoption through third-party vendors.

wages by 0.8%.[5]

AI also reduces wage inequality within and across occupations. Within occupations, a one-standard-deviation increase in AI exposure lowers the 90-to-10 wage ratio by about 2% by increasing employment of less skilled workers, consistent with experimental evidence that AI tools raise productivity more for lower-skill workers than for experts (Brynjolfsson et al. 2025, Peng et al. 2023, Kanazawa et al. 2022). Across occupations, it reduces average earnings and education more in initially high-wage jobs, compressing wage differences between occupations. Together, these results suggest that AI lowers skill barriers and opens access to roles previously held by more qualified workers (Autor 2024).

These average effects, however, mask important heterogeneity between office and production workers. AI raises employment in production-related occupations, such as agriculture and manufacturing, by about 5%. In contrast, it reduces employment in administrative occupations by the same amount. Employment among managers and high-skilled professionals remains largely unchanged. This pattern suggests that AI expands employment where workers use information to make decisions, coordinate production, or operate equipment, while reducing employment in occupations centered on administrative information processing. AI has the most negative effects in occupations that perform data processing tasks, such as data entry, form completion, and data collection. It has the most positive effects in occupations that perform decision making, supervise production teams, or operate machines.

Why does AI automate the office but expand the factory? The evidence points to a mechanism that depends on how occupations interact with information. AI increases employment in occupations that use information to make decisions, largely by drawing in younger and less educated workers. By contrast, it reduces employment in occupations centered on data processing, with especially large losses among young workers. These patterns are consistent with AI as a prediction technology (Agrawal et al. 2019b): it substitutes for workers whose main tasks are to process, organize, or record information, but complements workers who use information to make decisions, especially inexperienced workers who benefit from guidance.

Generative and predictive AI differ because they are used in different tasks. About 80 percent of generative AI registrations are management applications, which overlap most with administrative work. As a result, generative AI reduces employment on average, with larger losses among younger workers. Because predictive AI accounts for 99 percent of AI software

[5]According to the model by Autor and Thompson (2025), these results are consistent with AI reducing the expertise requirement of each occupation, bringing in more low-skilled workers and holding down average wages.

registrations, its positive effects in production dominate the aggregate results.

This paper contributes to the literature on the labor market effects of AI by introducing a new dataset and an instrumental variable strategy that identifies a novel channel: AI raises employment by boosting the productivity of less-experienced, lower-skilled workers. Much of the existing research emphasizes AI's labor-substituting role through task automation (Acemoglu and Restrepo 2018). Others argue that AI can augment workers and lead to job creation by increasing productivity (Agrawal et al. 2019b, Bresnahan 2021, Aghion et al. 2025). Agrawal et al. (2019b), Agrawal et al. (2019c), and Agrawal et al. (2019a) argue that AI provides a better prediction technology, which can increase demand for complementary tasks, such as judgment. Despite growing interest, empirical findings remain mixed. Some studies find little to no overall impact of AI on employment or wages, such as Humlum and Vestergaard (2025a) on AI chatbots, Acemoglu et al. (2022) using job postings, and Hampole et al. (2025) using resume data. Hampole et al. (2025), in particular, find that the productivity gains are larger than the displacement effect at firms adopting AI. Others, including Webb (2020), Aum and Shin (2024), and Aum and Shin (2025), show that AI adoption can reduce earnings and job opportunities. Hosseini and Lichtinger (2025) find that AI decreases the employment share of young workers, differently from standard software investment (Barth et al. 2023). In contrast, firm-level studies like Babina et al. (2024) and Adams et al. (2024) find positive effects of AI on firm growth and productivity, pointing to possible employment gains. Aghion et al. (2025), using firm-level data in France, find that AI increases employment overall but decreases it in administrative applications. Similarly, McElheran et al. (2025), Brynjolfsson and McElheran (2016) and Brynjolfsson et al. (2021), studying the adoption of AI, data-driven decision-making, and predictive technologies, find them to increase TFP and employment only when complementary investments are made, with losses of productivity in the short-run. Chalfin et al. (2016) show practical examples of how applying machine learning techniques to hiring can outperform standard hiring systems.

This paper helps reconcile the conflicting perspectives by showing that the labor market effects of AI vary across occupations. As Webb (2020) and Aum and Shin (2025) have argued, I find that AI reduces employment and wages among administrative workers, consistent with the displacement effects emphasized by Acemoglu and Restrepo (2018), Acemoglu et al. (2022), Hampole et al. (2025). At the same time, I find that AI increases employment in production occupations, supporting the view of Agrawal et al. (2019b) and Bresnahan (2021), who argue that AI can act as a productivity-enhancing tool that complements labor. Rather

than offering a single answer, my results suggest that the aggregate effect of AI depends on the composition of the labor force and how AI is used.

This paper contributes to the literature on how automation and technological progress affect labor markets. Seminal work by Graetz and Michaels (2018) and Acemoglu and Restrepo (2020) shows that the diffusion of industrial robots reduced low-skilled employment and wages. Building on this evidence, recent research moves to the firm level and shows that robot adoption changes employment and productivity within firms, with effects that can cumulate into sizable aggregate outcomes (Hubmer and Restrepo 2021, Martinez 2021). These effects are not uniformly negative: some studies find that adoption raises productivity enough to expand scale and increase employment at adopting firms (Jaravel et al. 2020, Humlum 2019). Automation also extends beyond industrial robots to other forms of capital and production technology (Boustan et al. 2022, de Souza and Li 2023). This broader adoption helps explain job polarization across countries, consistent with routine-biased technological change (Goos et al. 2014, Danieli 2026). Although automation can displace labor in adopting activities, reallocation and demand spillovers can offset part of these losses in other sectors (Autor and Salomons 2018). Technological progress can also generate new tasks that counterbalance displacement over time and decrease wage inequality in the long run (Autor et al. 2024, Kalyani et al. 2025).

I make two contributions to this literature. First, I study a new wave of technological progress and show that, unlike earlier waves of innovation, AI complements low-skilled workers and leads to lower inequality. Second, I broaden the measurement of automation technologies by using a dataset with detailed descriptions of software developed in Brazil. Most studies of automation focus on industrial robots (Humlum 2019, Jaravel et al. 2020), machines (Boustan et al. 2022, de Souza and Li 2023), or patents (Hémous et al. 2025). I show that AI software is also an automation technology for administrative workers.

The paper is organized as follows. Section 2 describes a simple model to characterize the empirical elasticities and the effects of AI on the labor market. Section 3 describes the data and the institutional context. Section 4 presents stylized facts about AI development in Brazil. Section 5 outlines the empirical strategy. Section 6 discusses the main results, and Section 7 addresses robustness. Section 8 concludes.

2 Theoretical Model

This section develops a simple task-based model to guide the empirical analysis. The model serves two purposes: it clarifies how to interpret the empirical findings and motivates an instrument based on shocks to AI development costs. Its central implication is that cheaper AI can either raise or reduce employment. The sign depends on two forces: how much lower AI prices expand task demand, and how easily firms substitute AI for workers within tasks.

2.1 Production and AI

Output and occupations. The final good combines output from broad occupational groups $g \in \mathcal{G}$ using a CES aggregator with elasticity of substitution $\chi > 0$. Within each broad group, a continuum of tasks produces output through another CES aggregator:

$$Y_g = \left(\int y(j)^{\frac{\psi-1}{\psi}} \, dj \right)^{\frac{\psi}{\psi-1}},$$

where Y_g is output of broad occupational group g, $y(j)$ is output of task j, and $\psi > 0$ is the elasticity of substitution across tasks within group g. Each broad group contains a finite set of occupations. The tasks in group g are partitioned across those occupations.

Worker heterogeneity. Within each occupation o, there are two types of workers: low-skilled and high-skilled workers, with wages $w_{o,L}$ and $w_{o,H}$. These types capture within-occupation heterogeneity in age, experience, education, or ability. For simplicity, both groups supply labor with the same elasticity η.

Task production. Consider a task j in the task set of occupation o. If that task is produced by workers of skill $s \in \{L, H\}$, output is

$$y_{o,s}(j) = A_{o,s}(j) \left[\ell_{o,s}(j)^{\frac{\sigma_{g,s}-1}{\sigma_{g,s}}} + K_{o,s}(j)^{\frac{\sigma_{g,s}-1}{\sigma_{g,s}}} \right]^{\frac{\sigma_{g,s}}{\sigma_{g,s}-1}}.$$

Here, $A_{o,s}(j)$ is a productivity shifter, $\ell_{o,s}(j)$ is labor input, and $K_{o,s}(j)$ is AI input.

Elasticity of substitution between AI and labor. The elasticity of substitution between AI and labor, $\sigma_{g,s}$, varies by broad group g and skill group s. This flexibility captures two patterns in the data. First, AI may complement low-skilled workers more than

high-skilled workers, which corresponds to $\sigma_{g,H} > \sigma_{g,L}$. Second, AI may have different labor-market effects in different broad occupational groups, such as office work and production work.

AI developers. AI software $K_{o,s}(j)$ is supplied competitively by software developers. Each AI developer combines services from different programmers to produce AI software:

$$K_o(j) = \prod_{l \in \mathcal{L}} x_{o,l}(j)^{\alpha_{j,l}}, \qquad \sum_{l \in \mathcal{L}} \alpha_{j,l} = 1,$$

where $x_{o,l}(j)$ is the quantity of programming services in language l used for task j, and $\alpha_{j,l}$ is the importance of language l for that task. If $c_{l,t}$ is the marginal cost of programming services in language l, the developer's cost-minimization problem implies the AI software price

$$q_o(j) = \prod_{l \in \mathcal{L}} \alpha_{j,l}^{-\alpha_{j,l}} c_l^{\alpha_{j,l}}$$

Equilibrium. An equilibrium consists of wages, $\{w_{o,L}, w_{o,H}\}_o$, and task-specific AI prices, $\{q_o(j)\}_j$, such that firms optimize, AI developers price software at unit cost, and product and occupation-level labor markets clear.

2.2 Effects of Cheaper AI on the Labor Market

Technological shock. Technological progress lowers AI prices heterogeneously across tasks:

$$\tau_o(j) = -d \ln q_o(j) \tag{1}$$

where a larger $\tau_o(j)$ denotes a larger decline in the AI price for task j in occupation o.

Many forces can lower AI costs. In the empirical analysis, I use variation from changes in the productivity of software programmers. In the model, this variation appears as a decline in c_l. The language-level cost shock is

$$\tau_l = -d \ln c_l, \tag{2}$$

$$\tau_o(j) = -d \ln q_o(j) = \sum_l \alpha_{j,l} \tau_l. \tag{3}$$

Heterogeneous exposure to technological shocks. Occupations differ in their exposure to the technological shock because their tasks rely on different language bundles. I define occupation o's AI shock exposure as the employment-weighted average decline in AI prices across its tasks:

$$\textit{AI Shock Exposure}_o = \int_{J_o} s_{M,o}(j)\lambda_o(j)\tau_o(j)\,dj = \sum_{l\in\mathcal{L}} \Omega_{o,l}\tau_l \tag{4}$$

where $s_{M,o}(j)$ is the AI expenditure share in task j and $\lambda_o(j) \equiv \ell_o(j)/L_o$ is the share of occupation-o employment allocated to task j. The second equality uses the language-level shocks in (2): occupation exposure depends on the size of each language shock, $\tau_{l,t}$, and on the occupation's dependence on that language, $\Omega_{o,l}$. Proposition 1 below explains how a decline in the price of AI affects employment across different occupations.[6]

Proposition 1 (AI has an ambiguous effect on employment) *Assume that the initial AI expenditure share is small. Then AI exposure affects employment in occupation o according to*

$$d\ln L_o \approx \underbrace{\frac{\eta}{\eta+\psi}\Xi_g}_{\textit{Scale Effect}} + \beta_o^L\ \textit{AI Shock Exposure}_o,$$

where Ξ_g is a function of parameters and

$$\beta_o^L = \frac{\eta}{\eta+\psi}\left[\underbrace{\psi}_{\textit{Demand Effect}} - \underbrace{\bar{\sigma_g}}_{\textit{Replacement Effect}}\right], \qquad \bar{\sigma}_g = \pi_o\sigma_{gL} + (1-\pi_o)\sigma_{gH}.$$

Proposition 1 highlights that the coefficient β^L captures two opposing forces. The first is a demand effect: when AI lowers task costs, demand for those tasks increases. This force is governed by ψ, the elasticity of substitution across tasks. The second is a replacement effect: as AI becomes cheaper, firms substitute workers for AI within a task. The strength of this effect is governed by the average elasticity of substitution between AI and workers, $\bar{\sigma}_g$. The elasticity β_o^L is positive when the demand effect dominates the replacement effect, and negative when the opposite is true.

[6]For tractability, I impose a balanced-exposure restriction: within each occupation, high- and low-skilled workers are equally exposed to the AI price shock. This restriction simplifies the expressions derived from the model while preserving the main economic mechanism.

AI can displace some occupations while expanding others. Proposition 1 implies that the sign of β_o^L may differ across occupations. AI reduces employment where substitution dominates and increases employment where demand expansion dominates. Therefore, the average effect of AI depends on the composition of the labor force, the source of identifying variation, and the AI applications measured. This logic helps explain why estimates can differ across studies. This intuition is formalized in Proposition 2.

Proposition 2 (Average Treatment Effect Depends on Sample Composition) *The linear projection of occupation employment on AI exposure can be approximated by*

$$d\ln L_o \approx \frac{\eta}{\eta+\psi}\Xi_g + \beta^L \textit{ AI Shock Exposure}_o + \epsilon_o,$$

where

$$\beta^L = \sum_o \omega_o \beta_o^L,$$

with $\pi_o, \omega_o \in (0,1)$, $\sum_o \omega_o = 1$, *and* ϵ_o *orthogonal to AI Shock Exposure*$_o$.

AI can increase the share of low-skilled workers within occupations. If AI complements low-skilled workers more than high-skilled ones, i.e., $\sigma_{g,H} > \sigma_{g,L}$, it should increase employment of low-skilled workers by more, leading to a shift in employment toward the low-skilled group. Proposition 3 formalizes this idea.

Proposition 3 (AI can change the composition of workers within occupations) *Assume that the initial AI expenditure share is small. Then, the effect of AI on the share of low-skilled workers is given by:*

$$d\textit{Shr. Low-Skilled}_o \approx \Theta_o\left(\sigma_{g,H} - \sigma_{g,L}\right) \textit{AI Shock Exposure}_o,$$

where $\Theta_o > 0$ *is a set of parameters.*

AI affects average wages through both direct wage effects and worker composition. Because AI affects the composition of workers, average wages might go down even if

employment increases. To formalize that, let $\bar{w}_o$ denote the average wage in occupation o. A first-order decomposition gives

$$d \ln \bar{w}_o \approx \vartheta_{oL} d \ln w_{oL} + \vartheta_{oH} d \ln w_{oH} + \frac{w_{oL} - w_{oH}}{\bar{w}_o} \, dShr.\ \textit{Low-Skilled}_o,$$

where ϑ_{os} is the wage-bill share of skill group s in occupation o. The first two terms capture the direct effect of AI on skill-specific wages. The last term captures the composition effect. Because low-skilled workers typically earn less than high-skilled workers, a rise in the low-skilled share can lower average wages even when AI raises productivity and employment.

3 Data

The model identifies four relevant objects to identify the effects of AI: a representative description of what AI software does, which occupations perform related tasks, which programming languages are used to develop each application, and cost-shocks to AI development. This section introduces the data used to observe these objects and explains why Brazil offers an unusually clear setting for studying the labor-market effects of AI.

The primary data source is the software registry maintained by the National Institute of Industrial Property (NIIP), which contains a broad sample of in-house, third-party, and commercially marketed software developed in Brazil (Barbosa et al. 2022). The registry is central to the empirical design because it reports each program's intended use, technical features, and programming languages. These fields allow me to identify AI software, link it to occupational tasks, and construct the cost-shock instrument motivated by the model.

The registry is unusually comprehensive for two reasons. First, Brazil has a mature and self-sufficient IT sector: it is the 11th largest in the world and has the third-largest pool of software developers (GitHub Staff 2024). Second, regulatory barriers limit the import share of computer services to about 9%, meaning most software deployed by Brazilian firms is locally developed and therefore captured in the NIIP database.[7]

I use the NIIP registry to construct occupation-level measures of AI exposure and link these measures to RAIS, an administrative dataset that tracks all formal jobs from 2003 to 2024 in narrowly defined 6-digit occupational codes, allowing precise measurement of employment and wage dynamics. Finally, I use Stack Overflow data to construct programming-

[7] I discuss in detail the import share of computer services in Appendix B.1.

language-specific shocks to AI development costs: Stack Overflow posts measure growth in the U.S. stock of AI programming knowledge that Brazilian developers can reuse, lowering the cost of AI software development (Mamykina et al. 2011, Treude and Robillard 2016, Wu et al. 2019, Ponzanelli et al. 2014).

3.1 Administrative Software Registry Data

3.1.1 Institutional Background

Brazilian software law provides special copyright protection for software. Since 1987, firms in Brazil have been able to register software with the National Institute of Industrial Property (NIIP) under a special copyright system that grants 50 years of protection.[8] The registration process is inexpensive, keeps the source code confidential, and provides legal protection against unauthorized copying, giving firms strong incentives to register commercially relevant software.

A software registration contains examiner-verified information about the software. To register software, firms must submit an electronic form with the firm's name, the programmers' names, a descriptive title of the software, and the programming languages. Importantly, firms also provide two classifications that describe the software's purpose and technical characteristics.

The first classification, called "application domain," identifies the sector or area where the software will be used. There are 225 specific application domains grouped into 35 broad categories. For example, *MapForest*, an AI program that forecasts weather and predicts land productivity, is classified under Agriculture. Each software entry includes at least one application domain, though most list multiple; on average, each program lists three. Table A4 in the Appendix lists all 35 broad application domains together with their descriptions.

The second classification, called "technical class," organizes software by its technical function. The system includes 97 specific technical classes divided into 18 broad groups, ranging from fundamental systems like Operating Systems and Network Controllers to specialized tools like Encryption Software and AI. *MapForest*, for instance, is classified under four technical classes: Artificial Intelligence, Simulation, Pattern Recognition, and Office

[8] In Portuguese, the NIIP is called *Instituto Nacional de Propriedade Industrial*. The first software law was Law number 7,646/1987. It was later updated by Law 9,609/1998. The law defines software as a "set of instructions or statements, written in proper language, to be used directly or indirectly by a computer to obtain a certain result."

Automation. Table A3, in Appendix B.1.2, lists the broad technical classification groups with their descriptions.

Software registration provides proof of ownership. The information provided by the firm is evaluated by an examiner at the NIIP and, if approved, the firm receives a registration certificate that serves as official legal proof of ownership. Figure 1 shows one such certificate. These certificates are commonly used in cease-and-desist letters, piracy takedown requests, and intellectual property lawsuits.[9]

Figure 1: **Example of a Software Registration**



FEDERATIVE REPUBLIC OF BRAZIL
MINISTRY OF THE ECONOMY
NATIONAL INSTITUTE OF INDUSTRIAL PROPERTY
DIRECTORATE OF PATENTS, COMPUTER PROGRAMS, AND TOPOGRAPHIES OF INTEGRATED CIRCUITS

Certificate of Software Registration

Case Nº: **BR512021001026-4**

The National Institute of Industrial Property issues this certificate of registration of a computer program, valid for 50 years starting from January 1st following the date of May 3, 2021, in accordance with §2, Article 2 of Law No. 9,609 of February 19, 1998.

Title: iMachine AI: Online Predictive Maintenance System for Electric Machines Using Machine Learning Techniques

Publication Date: 05/03/2021

Creation Date: 04/28/2021

Owner(s): TECHPLUS AUTOMACAO

Author(s): SAMARONE RUAS

Programming Language: PYTHON

Application Domain: AD-06; FQ-04; IF-10; IN-02

Technical Class: AP-01; AT-05; IA-01; IT-01; TC-03

Hash Algorithm: SHA-512

Digital Hash Summary:
cec9fe6467e2da2a9cc51fad00cfaedf52adc26dd5f352d4f4ed7c8600e399415383cc
2d542eab52cf2a49bf66955b20317e
bf21bafa7273d7edcdb534767588

Issued on: 05/25/2021

Approved by:
Carlos Alexandre
Fernandes Silva Head
of DIPTO

Notes: This figure shows a translated certificate of software registration. The application domains correspond to "factory planning", "metrology", "information management", and "manufacturing technology". The technical classes are "application", "industrial automation", "artificial intelligence", "instrumentation", and "pattern recognition".

[9]In the registration process, firms also provide a hash summary: a unique, fixed-length string generated by applying a cryptographic hash function to the source-code. Because any modification to the code would alter the hash, this digital "fingerprint" serves as proof that the firm possessed that exact version of the code on the filing date. The hash can later be compared to the confidential source code to establish authorship in legal proceedings.

Why are firms registering their software? Firms and developers both have concrete reasons to register software. For firms, registration helps protect intellectual property, participate in government procurement of software and IT services, document prior innovation when applying for R&D credit, and signal the value of intangible assets to investors (Torquato 2022, Associação Brasileira das Empresas de Software (ABES) 2024). For developers, it provides verifiable proof of experience that can be used on resumes and professional profiles. This credential carries substantial economic value: de Souza et al. (2024) documents that programmers receive a 5% permanent wage increase after being listed on a registered software project.

3.1.2 Administrative Software Registry Data

I extracted information from the universe of NIIP certificates of software registration between 1987 and 2024. For each entry, I observe all the information in the certificate of software registration (Figure 1). In the empirical section, I focus on AI software, defined as software with the technical class "Artificial Intelligence"("IA").

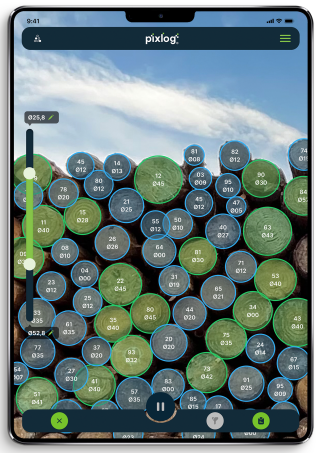
There is a total of 47,427 registrations, which, importantly, does not reflect the number of distinct software products. Registration is a legal tool, not a technical unit. A comparison of the software registrations with public online records reveals that firms usually bundle different software products in one registration. Legally, this approach offers the same level of protection as if the firm were to register each program separately. For the purposes of this paper, what matters is that commercially relevant software is typically registered at least once with a description of its intended use, even when firms bundle multiple products within a filing. I discuss the nuances of this database in further detail in the next section.

3.1.3 Selected Example

To illustrate the results that follow, it is useful to consider a concrete example from the database. Figure 2 shows Pixlog, an AI-based mobile application and management platform that automates log measurement and verifies whether shipments meet contractual specifications. Traditionally, workers measure log dimensions by hand and repeatedly compute volumes across large timber piles, a process that is slow and error-prone (Mais Floresta 2023). This task is especially difficult for less-educated workers, who commonly work in this sector. By embedding measurement expertise in the tool and providing accurate operational information, Pixlog enables production workers to complete these tasks faster and more con-

sistently, raising their productivity (Mais Floresta 2023). At the same time, the platform automates administrative functions such as inventory tracking and order control, reducing the need for manual recordkeeping.

Figure 2: **Example of AI from the Database**



Notes: Image from Pixlog (2026). Pixlog is a software registered in the NIIP.

3.1.4 Coverage, strengths, and limitations

This subsection evaluates the NIIP registry as a measure of AI software development. Registration is widespread among Brazilian software firms, and survey evidence supports its broad coverage. The registry's main advantage is that it captures actual software developed for use by firms, including tools that lack patent-worthy novelty and third-party vendor software missed by hiring data. I also discuss limitations related to scope and registration practices and explain why they are unlikely to drive the results.

Registration is nearly universal among software developers. In 2016, 92% of software development firms with more than two years of operation had registered at least one software program.[10,11] Weighted by employment, this figure rises to 99.4%, and among firms

[10]I call a firm a software development firm if it has Classificação Nacional de Atividades Econômicas (CNAE) 2002 codes 6201500 (custom software development) or 6202300 (development and licensing of customizable software), and has hired a computer engineer at least once in its life-cycle.

[11]I restrict the sample to firms with more than two years of operation because, as shown in Section C.4, software development typically requires more than two years.

with more than 20 employees it reaches 96%.[12] These numbers suggest that the NIIP registry captures a large share of commercially developed software created in Brazil.

Survey with private sector specialists. To validate the dataset and better understand the degree of selection, I conducted a survey with IT firms and intellectual property lawyers, described in more detail in Section B.1.1. The survey provides three main findings. First, software registration in the NIIP is widely known among both intellectual property lawyers and IT professionals, with the vast majority of respondents having direct experience with the registration process. Second, respondents confirm that most commercial software developed in Brazil has at least one version registered, with firms primarily motivated by intellectual property protection. Third, to the extent that selection exists, it is driven by small firms producing consumer-facing applications such as video games and mobile apps. These categories are unlikely to affect the labor market outcomes studied in this paper.

Registration titles validate the AI classification. Appendix Figure A14, in the appendix, shows that common title terms include prediction tasks, such as forecast, recommendation, classification, recognition, and detection, as well as AI methods such as machine learning, neural networks, and deep learning. These patterns suggest that NIIP software classified as AI captures the predictive technologies emphasized in the AI literature (Agrawal et al. 2019b).

Weaknesses. The NIIP database has two main limitations: it covers only domestically developed software and treats registrations as legal rather than technical units. Despite these limitations, to my knowledge, the NIIP database remains the only dataset with information on actual software deployed by firms, making it uniquely valuable for studying AI's labor market effects.

The registry covers only software created in Brazil, omitting software adopted locally but developed abroad. In practice, however, this omission appears modest. Imported computer services accounted for only 9% of domestic computer service consumption in 2015—less than half the OECD average of 19%—and fewer than 0.02% of IT firms were foreign-owned in

[12]This figure represents a conservative lower bound on overall coverage because not all of these firms have developed their software: (i) some firms might only provide technical support and do not write proprietary code, (ii) some firms are third-party developers whose software is registered by the client, and (iii) young firms are still completing their first product. As shown in Section C.4, software development takes more than two years for 50% of firms.

2016. Appendix B.1 discusses this issue in more detail.

A second limitation is that registrations are legal records, not product-level records. Because a filing protects the submitted source code, firms can register a bundle of related programs together rather than filing separately for each software product. The registry is therefore best interpreted as describing the main set of software developed by a firm, not as a count of individual products. This means the total count of registrations (47,427) understates the actual number of individual programs. To address this, the empirical analysis relies on the textual content of each registration rather than on the count of filings. Even when a registration bundles multiple programs, it records the technical features and intended applications of the submitted software, allowing me to identify the AI technologies firms develop and the occupations they are likely to affect.

3.2 Other Datasets

I also use three additional datasets. Labor market data comes from RAIS, and task descriptions from the Brazilian Ministry of Labor. To construct the instrument, I use data from Stack Overflow.

RAIS. I use the publicly available version of *"Relação Anual de Informações Sociais"* (RAIS), a matched employer-employee dataset covering all formal workers in Brazil. It reports each worker's occupation, earnings, hours, and other job details. I rely on this version because it extends coverage through 2024.[13] For privacy reasons, the public version excludes person identifiers, so I use a panel at the occupation–year level. Since 2003, RAIS has used the occupational classification system *Classificação Brasileira de Ocupações 2002* (CBO 2002), which includes detailed occupation descriptions. To avoid crosswalk issues with older classifications, I restrict the analysis to 2003–2024. I exclude from the sample IT workers, government employees, and military personnel.

Description of Occupational Tasks. To measure workers' exposure to AI, I use data from the Ministry of Labor, which links each six-digit occupation in the CBO 2002 classification to detailed descriptions of its tasks. Because the CBO is used in labor-court proceedings to determine employment benefits and labor union association, these descriptions are long-

[13]The longitudinal version tracks workers over time but only runs through 2016, missing important years of AI development.

form, legally precise, and highly specific.[14] The dataset covers 2,641 occupations, each with an average of 64 task statements, for a total of 49,660 unique tasks.

Stack Overflow. To build the instrument, I use data from Stack Overflow, a widely used question-and-answer platform for programmers. On this platform, programmers post questions seeking help from the community on specific technical problems. Each post includes standardized tags indicating the programming language and type of application involved, selected from a community-maintained list. I collect data on all Stack Overflow posts created between 2008 and 2024.[15]

4 Statistics of AI Software

The software registry data presents two relevant facts about AI software development in Brazil.[16] First, AI is used across a wide range of fields: while administrative uses such as report generation remain common, 40% of application-domain classifications target production settings, including manufacturing, agriculture, and logistics. Second, two applications dominate production use: manufacturing execution systems that coordinate workflow and quality-control tools that perform real-time visual inspections.

AI is used in management and production. When firms register software, they report its application domains, that is, the areas where the software is intended to be used. Figure 3a presents the ten most common broad application domains for AI software. Managerial domains such as information management and administration are the most common, but they account for only 27% of application-domain classifications. AI is also widely used in healthcare, manufacturing, and agriculture, which together account for 16.7% of the total.[17]

[14] For instance, in the states of Rio de Janeiro and Paraná minimum wages depend on the worker's CBO occupation.

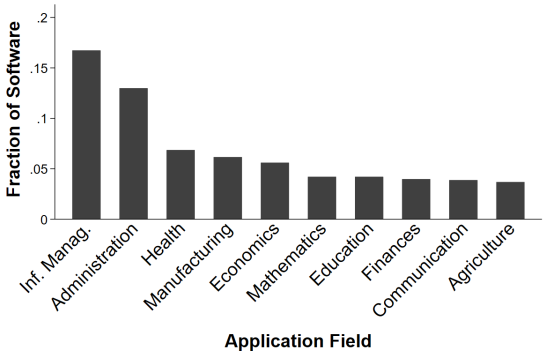
[15] B.3 provides a series of summary statistics for Stack Overflow posts.

[16] Additional statistics are reported in Appendix B.2.

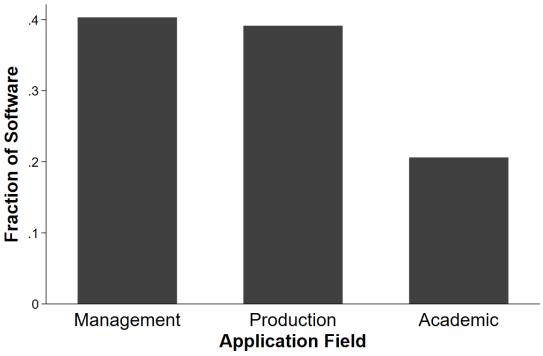
[17] Table A4 lists all application domains with descriptions.

Figure 3: **AI Software Registrations by Intended Use**

(a) Top 10 AI Application Domains



(b) AI on Management and Production



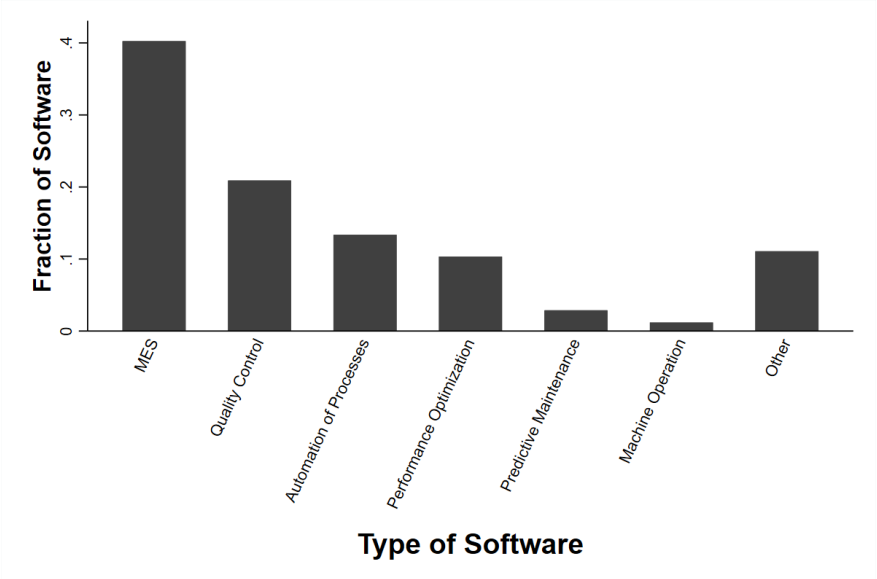
Notes: These figures present the distribution of AI software across different application domains, as defined by the NIIP. The application domain classification was developed by the NIIP to indicate the areas in which each software is intended to be used. Because a single software program may be assigned to multiple domains, each figure reports the fraction of each domain relative to the total number of domain classifications. Specifically, Figure 3a shows, for each domain, the ratio of the number of times that domain appears across all software to the total number of domain labels used. Figure 3b groups these domains into three broad categories. The "Management" category includes domains related to business administration, law, economics, finance, information storage, telecommunications, and human resources. The "Production" category includes agriculture, civil construction, ecology, education, energy, physics, chemistry, geography, geology, housing policy, hydrology, engineering, meteorology, soil studies, sanitation services, public infrastructure, transportation, biology, botany, and health. The "Academic" category covers domains in mathematics, cultural studies, politics, public policy, psychology, environmental policy, social studies, city planning, and social welfare.

Figure 3b groups all application domains into three categories: management, production, and academic. Management includes back-office functions such as human resources, accounting, and administration. Production includes non-management applied domains such as manufacturing, agriculture, and construction. Academic includes research-oriented applications in fields such as mathematics and the social sciences.[18]

Figure 3b shows that AI is split almost evenly between management and production applications, each accounting for about 40% of application domains. This balance challenges the prevailing view that AI primarily targets high-skill, white-collar occupations (Frey and Osborne 2017, Webb 2020, Felten et al. 2021). Because a comparable share of AI tools is used in production, blue-collar workers may be just as exposed to AI as administrative workers.

[18]AI software with application domains in business management, law, economics, finance, information storage, telecommunications, or human resources falls into the "Management" category. AI with application domains in agriculture, civil construction, ecology, education, energy, physics & chemistry, geography, geology, housing policy, hydrology, engineering, meteorology, soil studies, sanitation services, public infrastructure, transportation, biology, botany, and health is categorized as "Production." AI used in mathematics, cultural studies, politics, public policy, psychology, environmental policy, social studies, city planning, and social welfare is classified as "Academic."

Figure 4: **Specific Use of AI Software in Production**



Notes: This figure shows the distribution of AI production software across application categories. Each software entry was classified into one of the groups based on its title, technological class, and application domain by an LLM.

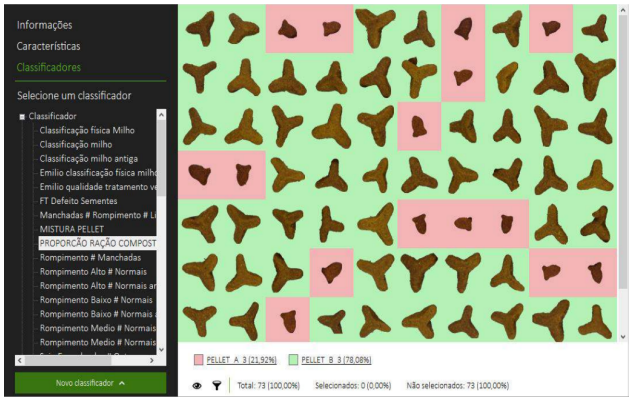
Production AI coordinates workflow, controls quality, and automates tasks. To better understand how AI is used in production, I classify each registration with a production-related application domain into one of seven mutually exclusive categories, based on the engineering literature and a manual review of program descriptions. Figure 4 plots the distribution of production AI software across the different categories.[19]

The most common production use of AI software is in Manufacturing Execution Systems (MES), which coordinate and control production processes. MES applications accounts for nearly 40% of all AI production applications in the database (Figure 4). One example is an AI system developed by Copel, a large electricity distribution company. Copel's system

[19]To construct the AI classifications, I began by reviewing the engineering and operations literature on the practical applications of AI in production. This review identified several recurring use cases, including Manufacturing Execution Systems (MES) (Chen et al. 2006, Capgemini 2020, Shojaeinasab et al. 2022), quality control (Ghelani et al. 2024, Tariq et al. 2024, Elsafty et al. 2024, Shi et al. 2023), industrial automation through collaborative robots and robotic arms (Borboni et al. 2023, Marr and Ward 2019), predictive maintenance (Keleko et al. 2022, Ucar et al. 2024, Murtaza et al. 2024), performance-optimization software that reduces electricity use or other input consumption (Evans and Gao 2016, Liu et al. 2022), and machine operation software, such as the vision-based picking aids used in Amazon warehouses (Marr and Ward 2019). Finally, I added a seventh category for software not well suited to any of the above categories. To classify each software entry, I used a large language model (LLM) with frontier prompting practices: I provided Gemini with a description of each category (Kumar et al. 2023, Hansen et al. 2024, Ruckdeschel 2025), hand-coded examples for guidance (Brown et al. 2020), and decision rules (Halterman and Keith 2025, Fonseca and Cohen 2024, Shinohara and Kawazoe 2026), then instructed it to assign each AI software record to one of the predefined groups, providing it with all the information available in each software registration.

monitors household-level energy use and uses AI to predict power outages.[20] When the system detects risk, it automatically dispatches maintenance and inspection teams with the location, diagnosis, and suggested repairs, increasing the productivity of field workers. According to Consoli (2025), the software has prevented 1,700 outages. More broadly, AI-enabled MES has been shown to reduce production time, increase yields (Chen et al. 2006), and improve machine utilization rates (Capgemini 2020), which could raise productivity and labor demand.[21]

Figure 5: **Example of AI Software Used in Quality Control**



Notes: Example of AI software from Tbit. Image from Tbit (2024).

The second most common use of AI is in quality control (Figure 4), where computer-vision software detects defects in real time. A notable example in the database is Tbit's image-recognition system, which identifies visual impurities in products such as seeds, flour, and animal feed. In one deployment at a dog food factory, the software flagged defective units for removal from the production line, as illustrated in Figure 5. It also relayed defect rates and quality metrics to the Manufacturing Execution System (MES), which used the data to recalibrate machines, adjust input mixes, and guide operator decisions. As documented in the engineering literature, integrating AI into quality control improves precision and productivity by reducing inspection time, minimizing human error, and lowering defect rates (Ghelani et al. 2024, Tariq et al. 2024, Shi et al. 2023).

[20]The use of AI in electricity distribution is widespread. Eight other utility companies in Brazil have developed their own AI tools. For a review of how AI is used in electricity distribution, see Forootan et al. (2022).

[21]See Shojaeinasab et al. (2022) for a review of AI in MES.

5 Empirics

This section describes the empirical strategy to estimate the effect of AI on the labor market. I first construct an occupation-level measure of AI exposure by computing the cosine similarity between task descriptions and AI software descriptions. To address differences in terminology, I apply latent semantic analysis, which maps synonyms and technical jargon onto shared semantic dimensions. I then use an instrumental variable strategy that exploits heterogeneous exposure to the growth of AI programming knowledge across programming languages in the United States. These knowledge shocks change the cost of developing AI software in Brazil and generate variation in AI exposure that is plausibly unrelated to Brazilian labor-market conditions.

5.1 AI Exposure

Text similarity between occupations and AI software. To measure AI exposure, I link the tasks performed by workers to those carried out by AI software. This measure captures the overlap between AI software and workers' tasks, which could arise from either potential substitution or complementarity between them. I construct this measure in three steps. First, I collect detailed task descriptions for each occupation from the Brazilian Ministry of Labor. Occupations are defined narrowly at the six-digit level, giving 2,641 distinct occupations in total.[22] On average, each occupation has 64 tasks, for a total of 49,660 unique tasks. Second, I construct descriptions for each AI software record by combining its title, the description of the listed application domains, and description of technological classes. Third, I use latent semantic analysis to calculate the cosine text similarity, $\rho_{o,s}$, between the tasks of occupation o and the description of AI software s. Appendix C.1 provides further details.

Exposure to AI. I define the exposure of occupation o to AI software available by year t as

$$AI\ Exposure_{o,t} = \sum_{s} \rho_{o,s} \mathbb{I}_s \left\{ t(s) \leq t \right\}, \tag{5}$$

[22] As an example of the narrowness of the occupational classification, economists in agriculture, finance, manufacturing, public finance, environmental, and regional economics each have their own entry.

where $\rho_{o,s}$ is the text similarity between occupation o and software s, and $\mathbb{I}_s\{t(s) \leq t\}$ equals one if software s was created in or before year t. Because firms register software shortly before commercialization, $t(s)$ also marks the year the software reaches the market. Thus, $AI\ Exposure_{o,t}$ measures occupation o's cumulative exposure to AI software commercialized by year t. Appendices C.2 and C.3 report summary statistics for this measure.

5.2 Empirical Model

Main empirical model. I use the following empirical model:

$$y_{o,t} = \beta AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t}, \tag{6}$$

where $y_{o,t}$ is a labor-market outcome, such as employment or wages, for occupation o in year t. $AI\ Exposure_{o,t}$ is defined in equation (5) and normalized to have mean zero and standard deviation one. Therefore, β is the effect of a one-standard-deviation increase in AI exposure. The terms μ_o and μ_t denote occupation and year fixed effects. Standard errors are clustered at the occupation level.[23]

Controls for omitted-variable bias. The control vector $X_{o,t}$ includes four variables addressing potential confounders. First, one-digit occupation-by-year fixed effects absorb shocks common to broad occupational groups. In the model derived in Section 2, they also capture the scale effect. Second, a U.S. patent-based AI exposure measure controls for exposure to foreign-developed AI, even though imported software accounts for less than 10% of the Brazilian market, as discussed in Section B.1.[24] Third, the lagged outcome, $y_{o,t-1}$, controls for differential trends. Fourth, a lagged non-AI software exposure measure controls for the adoption of other software. The robustness checks show that the results do not depend on these controls.

5.3 Instrument

Exogenous variation: changes in the cost of developing AI software. Building on the model-based AI shock exposure in Equation (4), I construct an instrument exploiting dif-

[23]Each occupation o is a narrow six-digit occupational code that often embeds sectoral requirements. For this reason, the baseline model does not use the sector–occupation specification in Webb (2020). In the robustness checks, I show that a specification closer to Webb (2020) delivers qualitatively similar results.

[24]To construct exposure to U.S. patents, I follow Webb (2020).

ferential exposure to AI-related knowledge across programming languages. The instrument exploits two shocks that raise programmer productivity. First, the size of a programming language community in the USA. Languages with large, active communities have more libraries, tutorials, forums, and reusable code, allowing Brazilian developers to adapt existing resources already developed elsewhere.[25] Second, the creation of Stack Overflow, which reduced search, debugging, and implementation costs (Mamykina et al. 2011, Treude and Robillard 2016, Wu et al. 2019, Ponzanelli et al. 2014).[26] As these developments happened in another country, they are likely orthogonal to Brazilian labor market conditions but reduced the cost of developing AI software.

These cost reductions affect occupations differently. Since programming languages are tailored to specific applications, these savings benefit some applications more than others (Nanz and Furia 2015, Rabah et al. 2010, Meyerovich and Rabkin 2013). For example, growth in the U.S. Matlab community makes AI engineering tools cheaper to build, while growth in HTML favors AI web applications. As language-specific resources expand in the United States, Brazilian developers can build some AI applications at lower cost. Occupations whose tasks overlap with those applications become relatively more exposed to AI software.

Example: COBOL for financial transactions and R for statistical analysis. To illustrate how the instrument works, consider two programming languages: COBOL and R. COBOL (Common Business-Oriented Language) was created in the 1950s for business applications and remains widely used in financial systems and government operations (Ovide 2022, Bailey 2024, Gershgorn 2020, Sullivan 2025). Over 80% of in-person transactions and 95% of ATM swipes still rely on COBOL-based systems (Hartman 2015). Despite this prevalence, COBOL's popularity has declined since its inception. According to the TIOBE Index, which measures the popularity of programming languages, COBOL dropped from a top-12 programming language in 2003 to a top-22 in 2023 (TIOBE 2025). This decline led to a shortage of experienced programmers, outdated documentation, difficulties integrating with modern languages, and fewer AI packages. Overall, these constraints make AI software development for the financial sector relatively more costly (Navot 2024).

[25]The availability of resources and libraries has been shown to reduce the time and cost of software development (Delorey et al. 2007, Krein et al. 2010, Tambad et al. 2020, Vijayaraghavan et al. 2022).

[26]Stack Overflow is a question and answer forum in English. It has been shown to provide fast answers to technical questions (Mamykina et al. 2011), contain knowledge missing from official documentation (Treude and Robillard 2016), and supplies code that developers use (Wu et al. 2019). More directly, Ponzanelli et al. (2014) show in a controlled experiment that Stack Overflow increased task completion.

Following a different trend from COBOL, R, created in the 1990s for statistical analysis, has experienced a remarkable increase in popularity. According to the TIOBE Index, R is the 15th most popular programming language in 2025 (TIOBE 2025). As a result, developing AI software for statistical analysis is comparatively less costly due to the large pool of available resources for R.

Therefore, because AI software is more costly to develop in COBOL than in R, occupations linked to banking automation, such as bank tellers and bank managers, become less exposed to AI than occupations linked to statistical reporting, such as archivists. The instrument identifies the effect of AI on employment by comparing employment growth across occupations differentially exposed to these language-specific cost shocks. The identifying assumption is that changes in U.S. AI programming resources for a given language are unrelated to labor-market trends in Brazilian occupations exposed to that language, except through their effect on AI development costs.

Instrument implementation. Using these insights and following the model-derived counterpart in Equation (4), the instrument is given by

$$\textit{AI Ease-of-Dev}_{o,t} = \frac{\sum_l \overline{\textit{AI Exposure}}_{l,o,t=0} \times \textit{Stack Overflow AI Posts}_{l,t}}{\sum_l \textit{Stack Overflow AI Posts}_{l,t}} \times \mathbb{I}\{t \geq 2008\} \quad (7)$$

where l indexes programming languages. $\overline{\textit{AI Exposure}}_{l,o,t=0}$ is the average text similarity between occupation o and software written in language l before 2003. This baseline similarity captures the specialization of programming languages: some occupations are more exposed to tasks commonly performed by software written in particular languages. $\textit{Stack Overflow AI Posts}_{l,t}$ is the number of Stack Overflow posts in language l that mention AI terms. This variable captures both the advantages of having access to knowledge of a broader community of experts and access to organized knowledge in Stack Overflow. The indicator $\mathbb{I}\{t \geq 2008\}$ is one after Stack Overflow was created in 2008.

Because the endogenous regressor is cumulative AI software exposure, I use the cumulative ease-of-development measure as the instrument:

$$\textit{AI Ease-of-Dev IV}_{o,t} = \sum_{k=2003}^{t} \textit{AI Ease-of-Dev}_{o,k}. \quad (8)$$

To control for the fact that Stack Overflow AI posts correlate with other Stack Overflow

posts, I also control for a non-AI ease-of-development measure constructed from non-AI Stack Overflow posts and occupation-level exposure to non-AI software.

Identifying Assumption. The instrument exploits differential exposure across occupations to an exogenous decline in the cost of developing AI software. The identifying assumption is a parallel-trends condition: in the absence of these cost shocks, the labor market outcomes of high- and low-exposure occupations would have followed similar trajectories (Goldsmith-Pinkham et al. 2020). To support this identifying assumption, in the next section I show that the instrument does not correlate with other technologies developed in Brazil, it is not associated with international technological progress in machines or equipment, it does not correlate with trade shocks, and is not correlated with labor market trends in Brazil.

5.4 Instrument Validation

To validate the identifying assumption, I show that the instrument is uncorrelated with other technological shocks, technological progress in tools and equipment, trade shocks, and pre-existing labor-market trends.

Economically irrelevant correlation with other technologies. Table A9 shows the results of a regression of the instrument on occupation-level exposure to non-AI software, patents, and industrial design. I construct these exposure measures with the same text-similarity approach used for AI exposure and normalize all variables to have standard deviation one. The instrument is uncorrelated with patents and industrial design. The correlation with non-AI software is statistically significant but economically small: a one-standard-deviation increase in the instrument raises non-AI software exposure by only 0.008 standard deviations. For comparison, the instrument leads to a 9,900% larger effect on AI software.

The instrument is not correlated with technological progress on tools. A concern is that changes in programming-language knowledge may coincide with technological progress in tools and equipment used by the same occupations. To test this channel, I link each machine, tool, and equipment category in the Harmonized System (HS) to occupations using the text similarity between product descriptions and occupational tasks. I then use these similarities as weights to compute occupation-level exposure to the import prices of

the products most related to each occupation. Table A10 shows no systematic correlation between the instrument and Brazilian import prices for capital goods, machines, or high-tech equipment.[27]

The instrument is not correlated with trade shocks. Another concern is that U.S. programming-language shocks may be correlated with sectoral shocks that affect Brazil through trade. I evaluate this concern by assigning trade exposure to occupations using their sectoral employment distribution. Table A11 shows no systematic correlation between the instrument and import prices, total imports, or export prices, suggesting that the instrument is not simply capturing sectoral shocks.

The instrument is not correlated with labor market trends. Table A12 tests whether the instrument predicts lagged labor-market outcomes. The estimates show no systematic relationship with lagged employment, education, demographic composition, wages, or hourly wages. Figure A17 shows that the instrument also does not correlate with lagged employment in different age groups. This evidence suggests that the instrument is not driven by pre-existing labor-market trends.

6 Results

In this section, I show that AI increases employment, with stronger effects among young, less-educated, inexperienced, and low-ability workers. This compositional shift lowers average wages and compresses wage inequality. Yet these average effects hide substantial heterogeneity: employment expands in production roles related to decision making, management, or machine operations; but contracts in administrative occupations related to data processing. The evidence points to a mechanism in which AI replaces tasks centered on processing information, while lowering skill barriers in tasks that use information to act or make decisions.

[27]This result does not imply that programming-language knowledge does not lead to cheaper tools and equipment. Instead, it implies that the tools that benefit from more AI knowledge are not the same as those used by workers more exposed to AI.

6.1 AI increases employment of low-skilled workers

AI ease-of-development instrument predicts AI exposure. Table 1 shows that the AI ease-of-development instrument strongly predicts AI exposure. In the baseline specification in Column 2, a one-standard-deviation increase in the instrument raises AI exposure by about 0.78 standard deviations. The F-statistic far exceeds 10, so the instrument comfortably passes the weak-instrument test.

Table 1: **First-Stage: Ease-of-Development Increases AI Exposure**

	(1) *AI Exposure*	(2) *AI Exposure*	(3) *AI Exposure*	(4) *AI Exposure*	(5) *AI Exposure*
AI Ease-of-Dev IV	0.823***	0.792***	0.932***	0.911***	0.910***
	(0.0322)	(0.0327)	(0.0392)	(0.0394)	(0.0405)
N	54,096	54,096	53,592	54,012	53,508
R^2	0.650	0.625	0.624	0.600	0.599
F	997.1	856.2	876.9	792.9	791.1
Fixed Effects Interacted with Year:					
1-Digit Occ.	-	✓	✓	-	-
Main Sector	-	-	✓	-	✓
2-Digits Occ.	-	-	-	✓	✓

Notes: This table presents the estimates of the first-stage. The instrument is given by Equation (8) and the endogenous variable is the exposure to AI, defined in Equation (5). All columns include occupation and year fixed effects. Column 2 adds year-by-1-digit-occupation fixed effects. Column 3 adds year-by-sector fixed effects to Column 2. Column 4 adds year-by-2-digit-occupation fixed effects. Column 5 combines year-by-sector and year-by-2-digit-occupation fixed effects. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Tables A13 and A14 in the Appendix clarify where the instrument gets its identifying variation. Table A13 shows that when a programming language receives more Stack Overflow posts, Brazilian developers write more software in that language. Table A14 shows that these effects are larger for technical classes that a programming language is more specialized in. Together, these patterns support the interpretation that U.S. shocks to resources for a programming language reduce development costs for specific AI applications, changing both the composition of software built and the set of occupations that face greater AI exposure.

AI leads to net employment gains. Table 2 reports the effect of AI on employment under different controls. AI exposure raises employment in every specification. In the baseline specification in Column 2, a one-standard-deviation increase in AI exposure raises employ-

ment by 3%. Table A15 in the Appendix decomposes this net gain: AI increases both hiring and separations, but hiring rises more, producing positive net job growth.

The positive employment effect is stable across specifications. Relative to the baseline in Column 2, Column 1 removes the year-by-one-digit-occupation fixed effect control. Column 3 adds year-by-sector fixed effects, where sector is the primary two-digit industry for each occupation.[28] Column 4 includes year-by-two-digit-occupation fixed effects, and Column 5 combines year-by-sector and year-by-two-digit-occupation fixed effects. The estimated coefficients are positive in all specifications.

Table 2: **AI Increases Employment**

	(1)	(2)	(3)	(4)	(5)
	log(Employment)	*log(Employment)*	*log(Employment)*	*log(Employment)*	*log(Employment)*
Panel A: 2SLS					
AI Exposure	0.0455***	0.0303***	0.0339***	0.0246**	0.0250**
	(0.00946)	(0.0101)	(0.0112)	(0.0113)	(0.0117)
Panel B: OLS					
AI Exposure	0.00814*	0.00509	0.00498	0.00244	0.00286
	(0.00444)	(0.00483)	(0.00504)	(0.00500)	(0.00520)
N	50,441	50,441	50,217	50,364	50,140
Fixed Effects Interacted with Year:					
1-Digit Occ.	-	✓	✓	-	-
Main Sector	-	-	✓	-	✓
2-Digits Occ.	-	-	-	✓	✓

Notes: This table reports estimates of the effect of AI on log employment, following the specification (6). Panel A presents the 2SLS estimates using the AI ease-of-development instrument defined in Equation (8). Panel B shows the corresponding OLS estimates. All columns include year and occupation fixed effects, lagged employment, exposure to non-AI software, exposure to U.S. AI patents, and the non-AI component of the ease-of-development instrument. Column 2 adds a year-by-1-digit-occupation fixed effect. Column 3 adds year-by-sector fixed effects, where sector is defined as the primary two-digit industry for each occupation. Column 4 adds year-by-2-digit-occupation fixed effects. Column 5 combines year-by-sector and year-by-2-digit-occupation fixed effects. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

The OLS estimate in Table 2 is small and not statistically different from zero. This pattern suggests that AI development in Brazil may target occupations already on declining employment trends, making the 2SLS strategy important for isolating supply-driven variation in AI exposure.

[28] I construct the year-by-sector fixed effect by first calculating, for each six-digit occupation, employment shares across two-digit National Classification of Economic Activities sectors. I then assign each occupation to the sector with the highest share and use this dominant sector to define the sector-by-year fixed effect. Because the occupational classification is narrow, the dominant sector accounts for 63 percent of employment on average.

AI shifts employment toward younger, less educated, and inexperienced workers. Table 3 shows that AI reallocates employment toward younger, less educated, and less experienced workers. Column 1 shows a decline in average age, Column 2 a decline in average years of education, and Column 3 a decline in average firm tenure. Table A16 in the Appendix shows that this compositional shift reflects larger employment gains for these demographic groups, rather than employment losses for other workers.

Figure 6 shows the effect of AI on employment of different age groups. AI has a significant effect on employment of workers aged between 20 and 39 years old, with insignificant effects for other age groups.

Table 3: **AI Increases Employment of Younger and Less Skilled Workers**

	(1) *log(Age)*	(2) *log(Yrs. Education)*	(3) *log(Tenure)*
2SLS			
AI Exposure	-0.00488^{***}	-0.00267^{*}	-0.0224^{***}
	(0.00150)	(0.00156)	(0.00807)
N	50,441	50,435	50,441

Notes: This table reports 2SLS estimates of the effect of AI exposure on worker composition. Column 1 reports the effect on log average age; Column 2 on log average years of education; and Column 3 on log average tenure. Standard errors appear in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

AI has no effect on average wages due to the entry of lower-ability workers. Column 1 of Table 4 shows that AI exposure has no significant effect on wages. To test whether this effect reflects composition rather than lower pay for existing workers, I decompose the effect into composition and price components. I estimate a Mincer regression relating individual wages to observable characteristics and worker fixed effects, then aggregate the worker fixed effects (which I call ability) and wage residuals to the occupation level.[29] Because ability is fixed at the individual level, changes in average ability capture shifts in workforce composition, while changes in the average wage residual capture wage changes unexplained by individual characteristics. To facilitate interpretation, I normalize both variables to have standard deviation of one.

This decomposition confirms the composition mechanism. Column 2 of Table 4 shows

[29]The ability measure uses the linked RAIS panel. Administrative lags in this dataset limit coverage to the period through 2016, so these results capture only the beginning of the AI boom. Appendix C.6 describes the Mincer regression and estimation procedure.

that AI lowers average ability, while Column 3 shows that AI raises wages conditional on worker characteristics, likely reflecting higher labor demand. The null effect on average wages therefore comes from compositional change: AI draws in workers who earn less based on both observable and unobservable characteristics. In the model presented in Section 2, this pattern is consistent with AI complementing low-ability workers.[30]

Table 4: **Effect of AI on Wages and Inequality**

	(1) *log(Wage)*	(2) *Ability*	(3) *Wage Residual*	(4) *log(P90/P10 Wage)*
2SLS				
AI Exposure	0.00107	-0.0301^{**}	0.132^{***}	-0.0209^{*}
	(0.00478)	(0.0136)	(0.0449)	(0.0115)
N	50,441	19,816	19,868	49,692

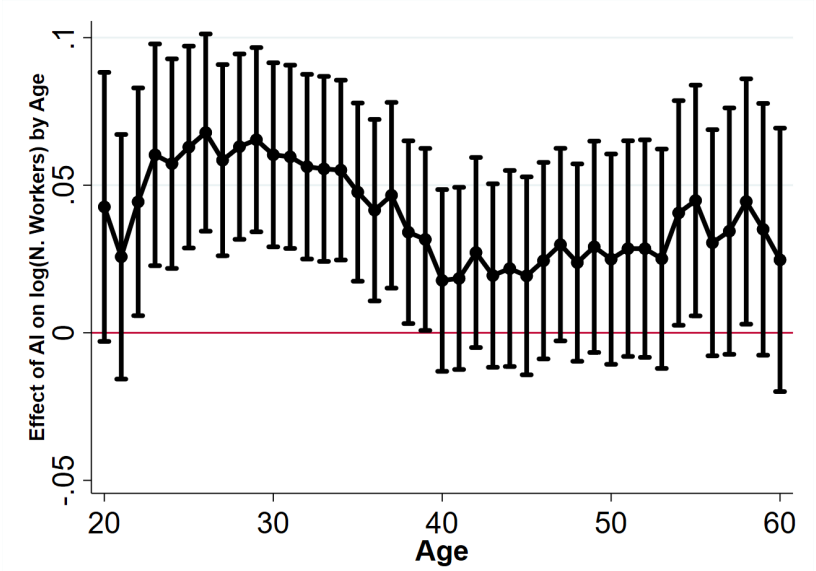
Notes: This table reports 2SLS estimates of the effect of AI on labor market outcomes, following the specification in equation (6). Each column uses the AI ease-of-development instrument defined in Equation (8). Column 1 reports the effect of AI on log average monthly earnings. Columns 2 and 3 show the effect of AI on the average ability and wage residuals, calculated using a Mincer regression at the occupation level and aggregated to occupations. Because ability and wage residuals require the panel structure of RAIS to be constructed, they are limited to the period from 2003 to 2016. Section C.6 explains in detail how these variables were calculated. In column 4 the variable of interest is the log p90-to-p10 ratio of wages within each occupation. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, exposure to U.S. AI patents, and the lagged outcome variable. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.

This increased demand for low-ability workers also compresses wage distributions within occupations. Column 4 of Table 4 shows that a one-standard-deviation increase in AI exposure lowers the ratio of the 90th decile to the 10th decile of wages by 2%. As AI raises relative demand for workers at the bottom of the wage distribution, it narrows wage gaps within occupations.

AI decreases inequality across occupations. The same composition mechanism that compresses wage gaps within occupations also reduces inequality across occupations. Table 5 shows that AI draws lower-ability workers by more into initially high-wage occupations, reducing wage differences across occupations.

[30]This conclusion is also supported by experimental evidence showing that AI leads to larger productivity gains for lower-skill workers than for experts (Kanazawa et al. 2022, Brynjolfsson et al. 2025, Gruber et al. 2020, Choi et al. 2023, Dell'Acqua et al. 2023, Noy and Zhang 2023, Peng et al. 2023).

Figure 6: **Effect of AI on Employment of Different Age Groups**



Notes: This figure reports the effect of AI on employment by worker age. Each point plots the estimated effect of AI on the log number of workers of a given age. All specifications use the AI ease-of-development instrument defined in Equation (8). Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, exposure to U.S. AI patents, and the lagged outcome variable. Standard errors are clustered at the occupation level. Vertical bars report 95% confidence intervals.

6.2 AI in the Office and the Factory

To understand the mechanism behind these results, I study heterogeneity in the effect of AI across occupations. The results show that AI reduces employment in administrative and data-processing tasks, but increases employment in production tasks that use information to guide decisions. This pattern supports a mechanism in which AI replaces information processing while complementing workers who use information to act.

AI replaces workers in the office but increases employment in the factory. Figure 7 estimates the effect of AI separately by one-digit occupational group.[31] AI raises employment in production-related occupations, such as workers in agriculture, extraction, manufacturing, maintenance and technicians, by around 5%. In contrast, it reduces employment in administrative occupations by 5%. Employment among managers and high-skilled professionals, such as engineers and chemists, remains largely unchanged.

[31]To create this figure, I allow the parameter of interest to vary by one-digit CBO occupational categories. The model is now given by $y_{o,t} = \beta_{O(o)} AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t}$, where $O(o)$ is the broad occupational class of occupation o.

Table 5: **AI Reduces Inequality Across Occupations**

	(1) *log(Employment)*	(2) *log(Yrs. Education)*	(3) *log(Wage)*
AI Exposure	−0.0518	0.0556***	0.197***
	(0.0398)	(0.00718)	(0.0265)
AI Exposure × *Initial Wage*	0.00991**	−0.00746***	−0.0254***
	(0.00468)	(0.000858)	(0.00325)
N	46,763	46,757	46,763

Notes: This table reports 2SLS estimates of the effect of AI exposure on labor market outcomes, allowing the effect to vary with an occupation's initial wage. Initial wage is log average monthly earnings in 2003. Column 1 reports the effect on log employment; Column 2 on log average years of education; and Column 3 on log average monthly earnings. AI exposure and its interaction with initial wage are instrumented using the AI ease-of-development instrument and its interaction with initial wage. Controls include occupation fixed effects, year fixed effects, one-digit-occupation-by-year fixed effects, exposure to U.S. AI patents, exposure to non-AI software, the non-AI component of the ease-of-development instrument, and the lagged outcome variable. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$.

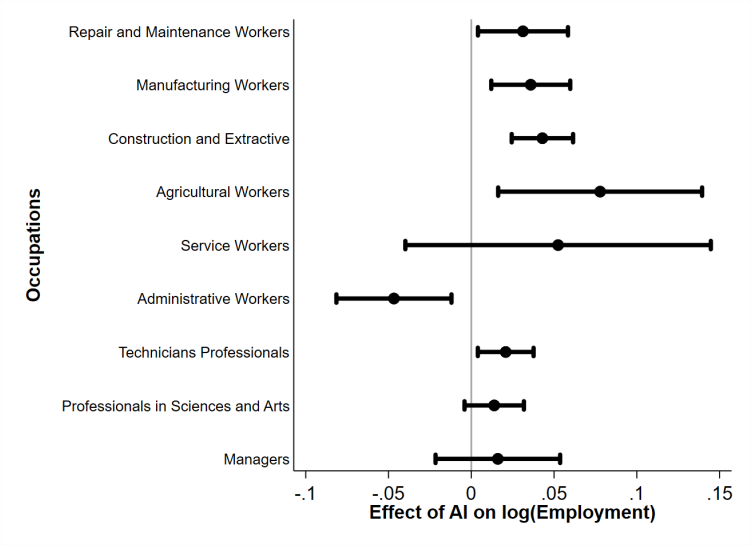
AI decreases employment of workers doing data processing tasks. To understand what workers in the occupations most affected by AI do, I classify the 49,660 task descriptions in the Brazilian occupational classification into 700 task groups using a semantic clustering procedure. Appendix C.7 describes this procedure. I then estimate whether the employment effect of AI differs across occupations that perform each task group. In particular, for each task group g, I run the following model:

$$y_{o,t} = \beta AI\ Exposure_{o,t} + \beta_g \mathbb{I}_o\left\{o\ performs\ task\ g\right\} \times AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t},$$

where $\mathbb{I}_o\left\{o\ performs\ task\ g\right\}$ equals one if occupation o contains at least one task in group g. I instrument AI exposure and its interaction with the task-group dummy using the AI ease-of-development instrument and its interaction with the same dummy.

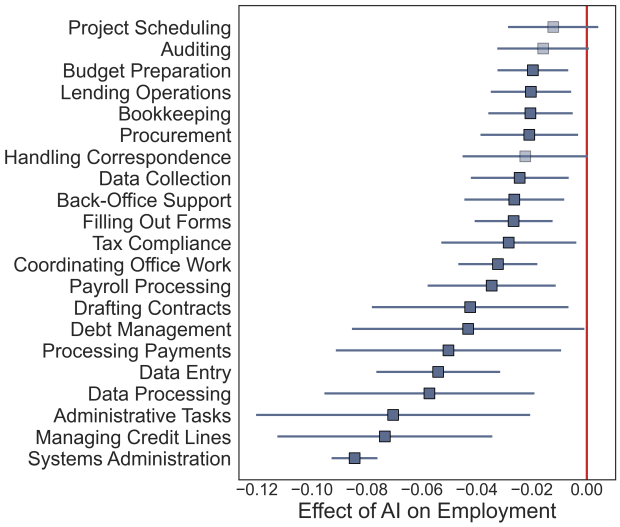
Figure 8 plots the differential effect of AI, β_g, for task groups with the largest employment losses: system administration, credit-line management, administrative tasks, data processing, data entry, form completion, tax compliance, and data collection. In occupations that perform these tasks, a one-standard-deviation increase in AI exposure reduces employment by 3 to 8 percentage points more than in occupations that do not. These results show that AI reduces employment most strongly in occupations centered on recording, processing, and organizing information.

Figure 7: **Differential Effect of AI on Employment Across Occupations**



Notes: This figure reports estimates of the effect of AI on employment for each one-digit occupational group. The model is: $y_{o,t} = \beta_{O(o)} AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t}$, where $O(o)$ denotes the broad occupational category of occupation o. All specifications use the AI ease-of-development instrument defined in Equation (8). In the first stage, the instrument is also interacted with a broad occupational group dummy. Standard errors are clustered at the occupation level. Bands represent 95% confidence intervals.

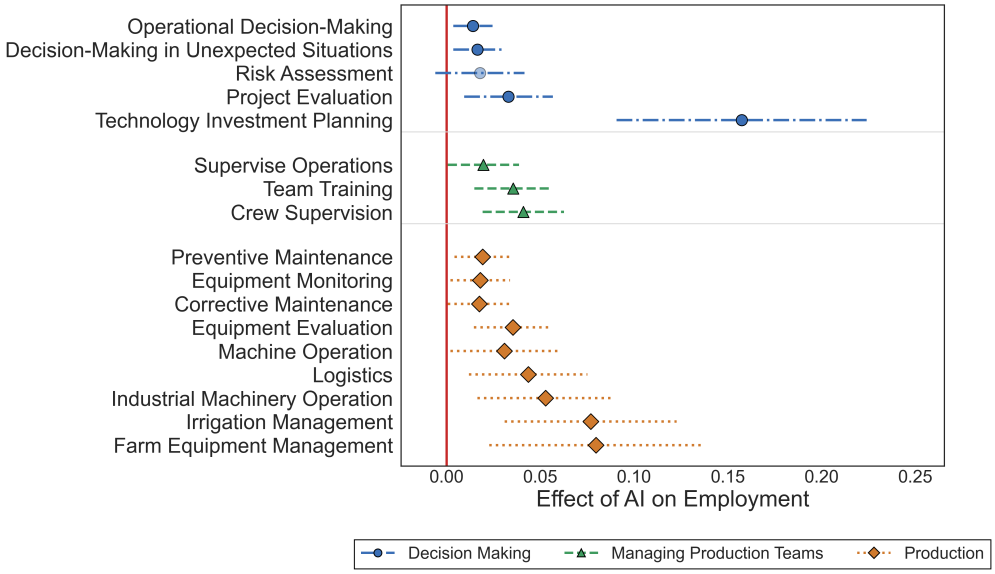
Figure 8: **AI Decreases Employment in Data Processing Tasks**



Notes: This figure reports task-group heterogeneity in the effect of AI exposure on log employment for selected administrative and data-processing task groups. Each point is the interaction coefficient β_g from $y_{o,t} = \beta AI\ Exposure_{o,t} + \beta_g AI\ Exposure_{o,t} \times D_{o,g} + \mu_o + \mu_t + \mu_{c(o),t} + \gamma y_{o,t-1} + \epsilon_{o,t}$, where $D_{o,g}$ equals one if occupation o contains at least one task in group g, and $\mu_{c(o),t}$ denotes one-digit-occupation-by-year fixed effects. AI exposure and its interaction with $D_{o,g}$ are instrumented using the AI ease-of-development instrument and its interaction with $D_{o,g}$. Controls include occupation fixed effects, year fixed effects, one-digit-occupation-by-year fixed effects, and lagged log employment. Standard errors are clustered at the occupation level. Horizontal bars report 95% confidence intervals.

AI increases employment in decision-making, production-supervision, and machine-operation tasks. In contrast, Figure 9 shows that the largest positive employment effects occur in task groups related to decision making, supervision of production teams, and machine operation. These are tasks in which better information helps workers respond to changing conditions, coordinate production, and operate equipment.

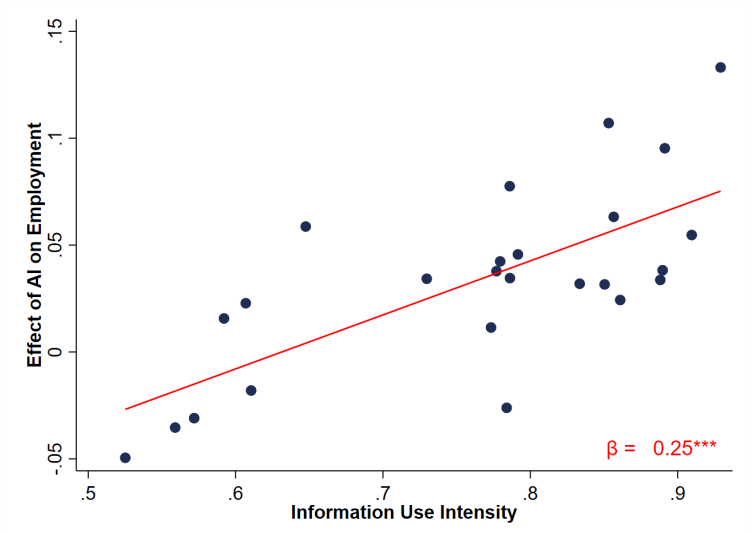
Figure 9: **AI Increases Employment in Decision Making, Management of Production Teams, and Machine Operation**



Notes: This figure reports task-group heterogeneity in the effect of AI exposure on log employment for selected task groups with positive employment effects. The task groups are classified into decision making, management of production teams, and machine operation using the semantic clustering procedure described in Appendix C.7. Each point is the interaction coefficient β_g from $y_{o,t} = \beta AI\ Exposure_{o,t} + \beta_g AI\ Exposure_{o,t} \times D_{o,g} + \mu_o + \mu_t + \mu_{c(o),t} + \gamma y_{o,t-1} + \epsilon_{o,t}$, where $D_{o,g}$ equals one if occupation o contains at least one task in group g, and $\mu_{c(o),t}$ denotes one-digit-occupation-by-year fixed effects. AI exposure and its interaction with $D_{o,g}$ are instrumented using the AI ease-of-development instrument and its interaction with $D_{o,g}$. Controls include occupation fixed effects, year fixed effects, one-digit-occupation-by-year fixed effects, and lagged log employment. Standard errors are clustered at the occupation level. Horizontal bars report 95% confidence intervals.

AI complements decision making and replaces data processing. These patterns are consistent with the view of AI as a prediction technology (Agrawal et al. 2019b). Better prediction complements workers who use operational information to decide what to do next, such as supervisors, maintenance technicians, and machine operators, but substitutes for labor in tasks that process, organize, or generate information. To test this mechanism directly, I follow de Souza and Li (2023) and classify tasks according to whether they primarily process data or use information to make decisions. I then define each occupation's information use intensity as the share of its tasks classified as information use. Higher values identify

Figure 10: **AI Automates Information Processing but Expands Information Use**



Notes: The figure plots the estimated effect of AI exposure on employment at the broad occupational level against an occupation's relative information use intensity. Each point corresponds to one of 20 broad occupational groups, narrower than the 1-digit level. The estimated effects are obtained from $y_{o,t} = \beta_{O(o)}\ AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t}$,where $O(o)$ denotes the broad occupational group containing occupation o. The relative information use intensity for broad occupation O is defined as $Relative\ Information\ Use\ Intensity_O = shr.\ of\ tasks\ that\ use\ information / shr.\ of\ tasks\ that\ use\ or\ process\ information.$

occupations that use operational information intensively; lower values identify occupations specialized in information processing and recordkeeping.[32]

Figure 10 shows the relationship between AI's employment effects and information use intensity. It plots the estimated effect of AI on employment against average information use intensity for each two-digit occupational group. The relationship is strongly positive: employment rises where workers use information to guide action, but falls where tasks center on processing data.

AI allows low-skilled workers to perform decision-making tasks. Table 6 estimates how the effects of AI on employment and worker composition vary with information use intensity. AI reduces employment in occupations intensive in data processing, but raises employment where workers use information intensively. The composition results show that these gains come from the entry of younger, less-educated, and lower-paid workers, reducing average age, education, and wages in those occupations. This pattern is consistent with AI raising the productivity of less-experienced workers by providing operational guidance in tasks that require decision making.

[32]Appendix C.8 describes in detail the classification process.

Table 6: **AI Has Larger Employment Effects in Information-Use Occupations**

	(1) *log(Employment)*	(2) *log(Age)*	(3) *log(Yrs. Education)*	(4) *log(Wage)*
AI Exposure	-0.0780^{***}	0.00252	0.00230	0.0153^{**}
	(0.0165)	(0.00244)	(0.00246)	(0.00754)
AI Exposure × *Inf. Use*	0.149^{***}	-0.0102^{***}	-0.00690^{**}	-0.0196^{**}
	(0.0209)	(0.00293)	(0.00322)	(0.00940)
N	50,364	50,364	50,358	50,364

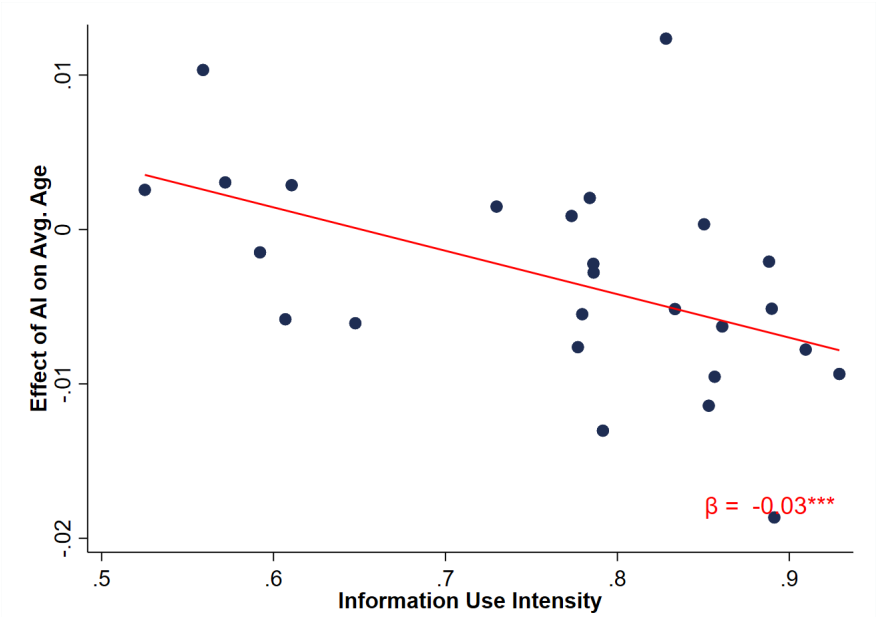
Notes: This table reports 2SLS estimates of the effect of AI exposure on labor market outcomes, allowing the effect to vary with an occupation's information use intensity. Information use intensity is the share of an occupation's tasks classified as closer to information use than to information processing. AI exposure and its interaction with information use intensity are instrumented using the AI ease-of-development instrument and its interaction with information use intensity. Column 1 reports the effect on log employment; Column 2 on log average age; Column 3 on log average years of education; and Column 4 on log average monthly earnings. Controls include occupation fixed effects, year fixed effects, one-digit-occupation-by-year fixed effects, exposure to U.S. AI patents, exposure to non-AI software, the non-AI component of the ease-of-development instrument, and the lagged outcome variable. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Figure 11 allows us to visualize the effect of AI on the composition of young workers in each occupation. It plots on the y-axis the effect of AI on age for every two-digits occupational groups against the average information use intensity in each of these occupations. AI leads to the entry of young workers among occupations intensive in data use for decision making but replaces them in occupations intensive in data processing.

Is generative AI different? Table 7 separates the effects of generative and predictive AI exposure. Predictive AI, which corresponds to 99% of AI software registrations, closely matches the average effects: it increases employment and reduces average age and experience. Generative AI has a negative effect on employment and increases average age and experience. Because 81% of generative AI software is used in management applications, generative AI has a stronger effect on replacing workers who perform data-processing tasks.

Discussion: AI shrinks the office and expands the factory. Taken together, the results show that AI affects administrative and production workers through different channels. It reduces employment in tasks organized around processing, recording, and transmitting information. At the same time, it expands employment in tasks that use information to make decisions, manage production, or operate machines. This mechanism helps reconcile the mixed predictions in the literature: whether AI replaces or complements workers depends on whether the occupation processes information or uses information to act.

Figure 11: **AI Brings in Younger Workers to Information-Use Intensive Occupations**



Notes: The figure plots the estimated effect of AI exposure on log average age at the broad occupational level against an occupation's relative information use intensity. Each point corresponds to one of 20 broad occupational groups, narrower than the 1-digit level. The estimated effects are obtained from $y_{o,t} = \beta_{O(o)}\ AI\ Exposure_{o,t} + \mu_o + \mu_t + X'_{o,t}\theta + \epsilon_{o,t}$,where $O(o)$ denotes the broad occupational group containing occupation o. The relative information use intensity for broad occupation O is defined as $Relative\ Information\ Use\ Intensity_O = shr.\ of\ tasks\ that\ use\ information / shr.\ of\ tasks\ that\ use\ or\ process\ information$.

7 Robustness

In Section 6, I showed that AI increases employment while shifting employment toward younger, less-educated, and less-experienced workers. Due to this compositional change, average wages are not affected by AI. I now show that these results are robust to controlling for U.S. labor-market shocks, estimating the model at the occupation-industry level, adding additional lags of the outcome, or removing controls.

Table 7: **Effects of Generative and Predictive AI on Worker Outcomes**

	(1) *log(Employment)*	(2) *log(Age)*	(3) *log(Yrs. Education)*	(4) *log(Experience)*	(5) *log(Wage)*
Gen. AI Exposure	-0.0119^{*}	0.00196^{***}	-0.00103	0.00940^{***}	0.00247
	(0.00618)	(0.000543)	(0.00171)	(0.00298)	(0.00213)
Pred. AI Exposure	0.0366^{***}	-0.00609^{***}	-0.00206	-0.0275^{***}	-0.000355
	(0.0111)	(0.00160)	(0.00191)	(0.00871)	(0.00525)
N	50,190	50,190	50,184	50,190	50,190

Notes: This table reports 2SLS estimates of the effects of generative and predictive AI exposure on worker outcomes. Column 1 reports log employment; Column 2 log average age; Column 3 log average years of education; Column 4 log average experience; and Column 5 log average wage. The two AI exposure measures are instrumented with the corresponding ease-of-development instruments for generative and predictive AI software. Controls include occupation fixed effects, year fixed effects, one-digit-occupation-by-year fixed effects, exposure to U.S. AI patents, exposure to non-AI software, the non-AI component of the ease-of-development instrument, and the lagged outcome variable. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10$,$^{**}p < 0.05$,$^{***}p < 0.01$.

Control for U.S. labor-market outcomes. One concern is that U.S. programming-language trends may capture global labor-market shocks that also affect Brazil. Panel A of Table 8 addresses this concern by controlling for the same outcome measured in the United States using the Current Population Survey. The results are unchanged: AI increases employment, lowers average age, education, tenure, within-occupation inequality, and doesn't affect average wages.

Occupation-industry regressions. Panel B re-estimates the baseline specification at the occupation-industry level, following Webb (2020). Occupations are defined using six-digit CBO codes, and industries are defined using four-digit CNAE2 codes. This specification is demanding because the Brazilian occupational classification is already narrow, creating many sparse occupation-industry cells. I include occupation-industry fixed effects and industry-year fixed effects to absorb permanent cell differences and sector-specific shocks. The estimates remain close to the baseline results.

Additional outcome lags. Panel C adds two lags of the dependent variable to absorb differential outcome dynamics across occupations. The estimates remain similar to the baseline: AI increases employment and shifts employment toward younger, less-educated, and less-experienced workers.

Regression without controls. Panel E removes all the controls exept for the occupation fixed effect and one-digit-occupation-year fixed effect. The estimates remain close to the baseline, showing that the main results are not driven by these controls.

Table 8: **Robustness: AI Increases Employment of Low-Skilled Workers**

	(1) *log(Employment)*	(2) *log(Age)*	(3) *log(Yrs. Education)*	(4) *log(Tenure)*	(5) *log(Wage)*	(6) *log(P90/P10 Wage)*
Panel A: Control for U.S. Labor Market Outcome						
AI Exposure	0.0302^{***}	-0.00488^{***}	-0.00264^{*}	-0.0224^{***}	0.000986	-0.0207^{*}
	(0.0101)	(0.00150)	(0.00156)	(0.00807)	(0.00478)	(0.0115)
Panel B: Occupation-Industry Level						
AI Exposure	0.0567^{***}	-0.00289^{***}	-0.00612^{***}	-0.00208	-0.00764^{***}	0.000249
	(0.00410)	(0.000773)	(0.000676)	(0.00364)	(0.00196)	(0.00375)
Panel C: Additional Outcome Lags						
AI Exposure	0.0262^{***}	-0.00380^{***}	-0.00270^{*}	-0.0180^{**}	0.00134	-0.0157
	(0.00969)	(0.00134)	(0.00151)	(0.00749)	(0.00438)	(0.0102)
Panel E: No Controls						
AI Exposure	0.0712^{***}	-0.00665^{***}	-0.00605^{***}	-0.0404^{***}	-0.0143^{***}	-0.0472^{***}
	(0.0155)	(0.00152)	(0.00145)	(0.00730)	(0.00503)	(0.00709)

Notes: This table reports robustness checks of the effect of AI exposure on labor-market outcomes. Each coefficient comes from a separate 2SLS regression using the AI ease-of-development instrument. Column 1 reports log employment; Column 2 log average age; Column 3 log average years of education; Column 4 log average tenure; Column 5 log average wage; and Column 6 the log p90-to-p10 wage ratio. Panel A controls for the U.S. version of each outcome; because the source output does not report tenure, Column 4 uses the baseline tenure estimate. Panel B estimates the model at the occupation-industry level and combines the occupation-industry outcome table with the corresponding p90-to-p10 wage-ratio table. Panel C includes additional lags of the dependent variable. Panel E excludes the lagged dependent variable and the U.S. AI exposure control. Unless otherwise noted, controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, exposure to U.S. AI patents, and the lagged outcome variable. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

8 Conclusion

The registry reports each program's intended use, technical characteristics, and programming languages, allowing me to observe a broad set of AI applications developed in Brazil. I use these records to construct occupation-level AI exposure by comparing software descriptions with occupational tasks. To address endogenous AI adoption, I instrument AI exposure with programming-language-specific shocks to AI development costs.

I find that, on average, AI increases employment without any effect on wages, driven by rising employment among younger, less-educated, and less-experienced workers. Within occupations, AI reduces wage inequality by benefiting less skilled workers. Across occupations,

however, the effects diverge: AI displaces administrative workers while expanding employment in production roles. These findings help reconcile competing views in the literature—confirming displacement effects in some settings while also presenting evidence that AI can complement labor and increase employment.

These results imply different policy responses across countries. Because AI displaces some workers while expanding employment for others, its net effect depends on a country's occupational mix. In economies with a large share of production workers, AI is likely to raise employment as workers move into production roles. In contrast, in economies dominated by administrative jobs, AI may reduce employment and force displaced workers to seek jobs in smaller production sectors or among professionals. Governments should therefore ease barriers to worker mobility across occupations, sectors, and regions. Where production sectors are large, this reallocation should mainly move workers from administrative jobs into production. Where agriculture and manufacturing are small, policy should instead support transitions from administrative work into professional roles.

This paper leaves several open questions for future research. Where do the low-skilled workers who gain employment through AI adoption come from? Given their age, AI may draw high-school–age workers into the labor market. Earlier entry could affect human capital accumulation and long-run growth. Another open question concerns administrative workers displaced by AI. What becomes of these workers, and how likely are they to transition into jobs that use AI? Answering these questions is essential for policy design and for understanding the long-run effects of AI.

References

Acemoglu, D., D. Autor, J. Hazell, and P. Restrepo (2022): "Artificial Intelligence and Jobs: Evidence from Online Vacancies," *Journal of Labor Economics*, 40, 293–340.

Acemoglu, D. and P. Restrepo (2018): "Artificial Intelligence, Automation, and Work," in *The Economics of Artificial Intelligence: An Agenda*, National Bureau of Economic Research, Inc, NBER Chapters, 197–236.

——— (2019): "The Wrong Kind of AI? Artificial Intelligence and the Future of Labor Demand," Working Paper w25682, National Bureau of Economic Research, Cambridge, MA, received May 30, 2019; accepted Nov 18, 2019.

——— (2020): "Robots and Jobs: Evidence from US Labor Markets," *Journal of Political Economy*, 128.

Adams, J., M. Fang, Z. Liu, and Y. Wang (2024): "The Rise of AI Pricing: Trends, Driving Forces, and Implications for Firm Performance," Working Paper Series 2024-33, Federal Reserve Bank of San Francisco.

Aghion, P., S. Bunel, X. Jaravel, T. Mikaelsen, A. Roulet, and J. Søgaard (2025): "How Different Uses of AI Shape Labor Demand: Evidence from France," *AEA Papers and Proceedings*, 115, 62–67.

Agoro, H. (2025): "Reducing Downtime in Production Lines Through Proactive Maintenance Strategies," .

Agrawal, A., J. Gans, and A. Goldfarb (2019a): "Prediction, Judgment, and Complexity: A Theory of Decision-Making and Artificial Intelligence," in *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press.

Agrawal, A., J. S. Gans, and A. Goldfarb (2019b): "Artificial Intelligence: The Ambiguous Labor Market Impact of Automating Prediction," NBER Working Papers 25619, National Bureau of Economic Research, Inc.

——— (2019c): "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Information Economics and Policy*, 47, 1–6, the Economics of Artificial Intelligence and Machine Learning.

Alam, M. F., A. Lentsch, N. Yu, S. Barmack, S. Kim, D. Acemoglu, J. Hart, S. Johnson, and F. Ahmed (2024): "From Automation to Augmentation: Redefining Engineering Design and Manufacturing in the Age of NextGen-AI," Working paper, Massachusetts Institute of Technology, Cambridge, MA.

Alaveras, G. and B. Martens (2015): "International Trade in Online Services," JRC Working Papers on Digital Economy 2015-08, Joint Research Centre.

Aoun, M. R. E., L. N. Tidjon, B. Rombaut, F. Khomh, and A. E. Hassan (2022): "An Empirical Study of Library Usage and Dependency in Deep Learning Frameworks," *arXiv preprint*.

Arora, A. and A. Gambardella (2005): "The Globalization of the Software Industry: Perspectives and Opportunities for Developed and Developing Countries," in *Innovation Policy and the Economy, Volume 5*, National Bureau of Economic Research, Inc, NBER Chapters, 1–32.

Associação Brasileira das Empresas de Software (ABES) (2024): "Por que registrar meu software?" Accessed: 2024-11-14.

Aum, S. and Y. Shin (2024): "Is Software Eating the World?" Working Paper 32591, National Bureau of Economic Research.

——— (2025): "The Labor Market Impact of Digital Technologies," *Federal Reserve Bank of St. Louis Review*, 107, 85–108, accessed: 2025-05-29.

Autor, D. (2024): "Applying AI to Rebuild Middle Class Jobs," Working Paper 32140, National Bureau of Economic Research.

Autor, D., C. Chin, A. Salomons, and B. Seegmiller (2024): "New Frontiers: The Origins and Content of New Work, 1940–2018*," *The Quarterly Journal of Economics*, 139, 1399–1465.

Autor, D. and A. Salomons (2018): "Is Automation Labor Share–Displacing? Productivity Growth, Employment, and the Labor Share," *Brookings Papers on Economic Activity*, 1–63.

Autor, D. and N. Thompson (2025): "Expertise," NBER Working Papers 33941, National Bureau of Economic Research, Inc.

Babina, T., A. Fedyk, A. He, and J. Hodson (2024): "Artificial intelligence, firm growth, and product innovation," *Journal of Financial Economics*, 151, 103745.

Bailey, C. (2024): "COBOL is a Cockroach: Why We Won't Get Rid of It Anytime Soon," `https://www.faircom.com/learn/blog/cobol-is-a-cockroach-why-we-wont-get-rid-of-it-anytime-soon`, accessed: 16 Feb 2025.

Barbosa, F., A. Souza, A. Souza, P. Nascimento, and A. Baganha (2022): "Análise Exploratória dos Registros de Software do Instituto Nacional da Propriedade Industrial (INPI) de 2018 a 2020," *Conjecturas*, 22, 765–777.

Barth, E., J. C. Davis, R. B. Freeman, and K. McElheran (2023): "Twisting the demand curve: Digitalization and the older workforce," *Journal of Econometrics*, 233, 443–467.

Benhanifia, A., Z. B. Cheikh, P. M. Oliveira, A. Valente, and J. Lima (2025): "Systematic review of predictive maintenance practices in the manufacturing sector," *Intelligent Systems with Applications*, 26, 200501.

Benz, S. and A. Jaax (2022): "The costs of regulatory barriers to trade in services: New estimates of ad valorem tariff equivalents," *Economics Letters*, 212, 110057.

Borboni, A., K. V. V. Reddy, I. Elamvazuthi, M. S. Al-Quraishi, E. Natarajan, and S. S. A. Ali (2023): "The Expanding Role of Artificial Intelligence in Collaborative Robots for Industrial Applications: A Systematic Review of Recent Works," *Machines*, 11, 111.

Boustan, L. P., J. Choi, and D. Clingingsmith (2022): "Automation After the Assembly Line: Computerized Machine Tools, Employment and Productivity in the United States," Working Paper 30400, National Bureau of Economic Research.

Bresnahan, T. (2021): *Artificial Intelligence Technologies and Aggregate Growth Prospects*, Cambridge University Press, 132–170.

Brown, T. B., B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. M. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei (2020): "Language Models are Few-Shot Learners," .

Brynjolfsson, E., W. Jin, and K. McElheran (2021): "The power of prediction: predictive analytics, workplace complements, and business performance," *Business Economics*, 56, 217–239.

Brynjolfsson, E., D. Li, and L. Raymond (2025): "Generative AI at Work*," *The Quarterly Journal of Economics*, 140, 889–942.

Brynjolfsson, E. and K. McElheran (2016): "The Rapid Adoption of Data-Driven Decision-Making," *American Economic Review*, 106, 133–39.

——— (2019): "Data in Action: Data-Driven Decision Making and Predictive Analytics in U.S. Manufacturing," Working Paper 3422397, Rotman School of Management, University of Toronto, date written: July 6, 2019.

Campello, R. J. G. B., D. Moulavi, and J. Sander (2013): "Density-Based Clustering Based on Hierarchical Density Estimates," in *Advances in Knowledge Discovery and Data Mining*, Springer, 160–172.

Campello, R. J. G. B., D. Moulavi, A. Zimek, and J. Sander (2015): "Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection," *ACM Transactions on Knowledge Discovery from Data*, 10, 1–51.

Capgemini (2020): "Artificial Intelligence (AI)-Driven Smart Factory Solution Provides Operators and Engineers with a New Level of Insight and the Ability to Adjust Production at a Moment's Notice," Tech. rep.

Chalfin, A., O. Danieli, A. Hillis, Z. Jelveh, M. Luca, J. Ludwig, and S. Mullainathan (2016): "Productivity and Selection of Human Capital with Machine Learning," *American Economic Review*, 106, 124–27.

Chen, R.-s., Y.-s. Tsai, and C.-c. Chang (2006): "Design and Implementation of an Intelligent Manufacturing Execution System for Semiconductor Manufacturing Industry," in *2006 IEEE International Symposium on Industrial Electronics*, vol. 4, 2948–2953.

Choi, J. H., D. Schwarcz, and K. E. Yeh (2023): "AI Assistance in Legal Analysis: An Empirical Study," Legal Studies Research Paper 23-22, University of Minnesota Law School.

Consoli (2025): "Inteligência artificial e termovisão reforçam eficiência da Copel e reduzem quedas de energia no Paraná," `https://aerp.org.br/redeaerp/inteligencia-artificial-e-termovisao-reforcam-eficiencia-da-copel-e-reduzem-quedas-de-energia-no-parana/`, publicado por Redação, com texto de Flávia Consoli — Associação das Emissoras de Radiodifusão do Paraná.

Danieli, O. (2026): "Skill-Replacing Technology and Bottom-Half Inequality," Working paper, Tel Aviv University; CEPR; IZA, previously titled "Revisiting U.S. Wage Inequality at the Bottom 50%".

de Souza, G., J. S. Herbstman, and J. Mannion (2024): "How Demand for New Skills Affects Wage Inequality: The Case of Software Programmers," Working Paper Series WP 2024-19, Federal Reserve Bank of Chicago.

de Souza, G. and H. Li (2023): "Robots, Tools, and Jobs: Evidence from Brazilian Labor Markets," Working Paper Series WP 2023-42, Federal Reserve Bank of Chicago.

Dell'Acqua, F., E. McFowland III, E. Mollick, H. Lifshitz-Assaf, K. C. Kellogg, S. Rajendran, L. Krayer, F. Candelon, and K. R. Lakhani (2023): "Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality," Working Paper 24-013, Harvard Business School, Boston, MA.

Delorey, D., C. Knutson, and S. Chun (2007): "Do Programming Languages Affect Productivity? A Case Study Using Data from Open Source Projects," 8–8.

Elsafty, A., M. Shaaban, A. El-Naggar, M. Okasha, and M. Elnaggar (2024): "AI-Driven Quality Management System for Sustainable Construction Projects: A Decision Support Model," *Energy Efficiency*.

Evans, R. and J. Gao (2016): "DeepMind AI Reduces Google Data Centre Cooling Bill by 40%," Accessed: 2025-05-08.

Ezell, S. J., R. D. Atkinson, and M. A. Wein (2013): "Localization Barriers to Trade: Threat to the Global Innovation Economy," Technical Report –, Information Technology & Innovation Foundation, Washington, DC, available at `https://www2.itif.org/2013-localization-barriers-to-trade.pdf`.

Felten, E., M. Raj, and R. Seamans (2021): "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal*, 42, 2195–2217.

Felten, E. W., M. Raj, and R. Seamans (2018): "A Method to Link Advances in Artificial Intelligence to Occupational Abilities," *AEA Papers and Proceedings*, 108, 54–57.

Fonseca, M. and S. B. Cohen (2024): "Can Large Language Models Follow Concept Annotation Guidelines? A Case Study on Scientific and Financial Domains," .

Forootan, M. M., I. Larki, R. Zahedi, and A. Ahmadi (2022): "Machine Learning and Deep Learning in Energy Systems: A Review," *Sustainability*, 14, 4832.

Frey, C. B. and M. A. Osborne (2017): "The future of employment: How susceptible are jobs to computerisation?" *Technological Forecasting and Social Change*, 114, 254–280.

Gershgorn, D. (2020): "Our Government Runs on a 60-Year-Old Coding Language — And Now It's Falling Apart," `https://onezero.medium.com/our-government-runs-on-a-60-year-old-coding-language-and-now-its-falling-apart-61ec0bc8e121`, accessed: 16 Feb 2025.

Ghelani, H. et al. (2024): "AI-Driven Quality Control in PCB Manufacturing: Enhancing Production Efficiency and Precision," *ResearchGate Preprint*, accessed: 2025-05-07.

GitHub Staff (2024): "Octoverse: AI leads Python to top language as the number of global developers surges," GitHub Blog.

Gmyrek, P., J. Berg, and D. Bescond (2023): "Generative AI and jobs a global analysis of potential effects on job quantity and quality," Tech. rep.

Goldsmith-Pinkham, P., I. Sorkin, and H. Swift (2020): "Bartik Instruments: What, When, Why, and How," *American Economic Review*, 110, 2586–2624.

Goos, M., A. Manning, and A. Salomons (2014): "Explaining Job Polarization: Routine-Biased Technological Change and Offshoring," *American Economic Review*, 104, 2509–2526.

Graetz, G. and G. Michaels (2018): "Robots at Work," *The Review of Economics and Statistics*, 100, 753–768.

Grootendorst, M. (2022): "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," .

Gruber, J., B. R. Handel, S. H. Kina, and J. T. Kolstad (2020): "Managing Intelligence: Skilled Experts and AI in Markets for Complex Products," NBER Working Papers 27038, National Bureau of Economic Research, Inc.

Halterman, A. and K. A. Keith (2025): "Codebook LLMs: Evaluating LLMs as Measurement Tools for Political Science Concepts," *Political Analysis*, 1–17.

Hampole, M., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2025): "Artificial Intelligence and the Labor Market," NBER Working Papers 33509, National Bureau of Economic Research, Inc.

Hansen et al. (2024): "Just Read the Codebook: Using Large Language Models for Sociopolitical Text Classification with Expert Annotations," .

Harms, P. and D. Shuvalova (2020): "Cultural distance and international trade in services: A disaggregate view," *Economic Systems*, 44, 100786.

Hartman, T. (2015): "COBOL Blues," `https://www.reuters.com/graphics/USA-BANKS-COBOL/010040KH18J/`, accessed: 16 Feb 2025.

Hémous, D., M. Olsen, C. Zanella, and A. Dechezleprêtre (2025): "Induced Automation Innovation: Evidence from Firm-Level Patent Data," *Journal of Political Economy*, 133, 1975–2028.

Hosseini, S. M. and G. Lichtinger (2025): "Generative AI as Seniority-Biased Technological Change: Evidence from U.S. Résumé and Job Posting Data," SSRN Working Paper No. 5425555.

Hubmer, J. and P. Restrepo (2021): "Not a Typical Firm: The Joint Dynamics of Firms, Labor Shares, and Capital–Labor Substitution," Working Paper 28579, National Bureau of Economic Research.

Hui, X., O. Reshef, and L. Zhou (2024): "The Short-Term Effects of Generative Artificial Intelligence on Employment: Evidence from an Online Labor Market," *Organization Science*, earlier version: CESifo Working Paper No. 10601.

Humlum, A. (2019): "Robot Adoption and Labor Market Dynamics," Job Market Paper, Princeton University.

Humlum, A. and E. Vestergaard (2025a): "Large Language Models, Small Labor Market Effects," Working Paper 33777, National Bureau of Economic Research.

——— (2025b): "The unequal adoption of ChatGPT exacerbates existing inequalities among workers," *Proceedings of the National Academy of Sciences*, 122, 2414972121–.

Ide, E. and E. Talamàs (2025): "Artificial Intelligence in the Knowledge Economy," *Journal of Political Economy*.

Instituto Nacional da Propriedade Industrial (INPI) (2015a): *Campo de aplicação*, Instituto Nacional da Propriedade Industrial, atualizado em 09/04/2015.

——— (2015b): *Tipos de programas de computador*, Instituto Nacional da Propriedade Industrial, atualizado em 09/04/2015.

Jaravel, X., P. Aghion, C. Antonin, and S. Bunel (2020): "What Are the Labor and Product Market Effects of Automation? New Evidence from France," CEPR Discussion Papers 14443, C.E.P.R. Discussion Papers.

Junqueira Botelho, A. J., G. Stefanuto, and F. Veloso (2005): "The Brazilian Software Industry," in *From Underdogs to Tigers: The Rise and Growth of the Software Industry in Brazil, China, India, Ireland, and Israel*, Oxford University Press.

Kalyani, A., T. Hassan, and P. Restrepo (2025): "New Technologies and Inequality," Working paper, Federal Reserve Bank of St. Louis, available at author's website.

Kanazawa, K., D. Kawaguchi, H. Shigeoka, and Y. Watanabe (2022): "AI, Skill, and Productivity: The Case of Taxi Drivers," CIRJE F-Series CIRJE-F-1202, CIRJE, Faculty of Economics, University of Tokyo.

Keleko, A. T., B. Kamsu-Foguem, R. H. Ngouna, and A. Tongne (2022): "Artificial intelligence and real-time predictive maintenance in industry 4.0: a bibliometric analysis," *AI and Ethics*, 2, 553–577.

Kogan, L., D. Papanikolaou, L. D. Schmidt, and B. Seegmiller (2023): "Technology and Labor Displacement: Evidence from Linking Patents with Worker-Level Data," Working Paper 31846, National Bureau of Economic Research.

Krein, J., A. Maclean, C. Knutson, D. Delorey, and D. Eggett (2010): "Impact of Programming Language Fragmentation on Developer Productivity: A Sourceforge Empirical Study," *IJOSSP*, 2, 41–61.

Kumar, S., C. Y. Park, and Y. Tsvetkov (2023): "Gen-Z: Generative Zero-Shot Text Classification with Contextualized Label Descriptions," .

Li, Y., X. Dang, H. Tian, T. Sun, Z. Wang, L. Ma, J. Klein, and T. F. Bissyandé (2024): "An Empirical Study of AI Techniques in Mobile Applications," *arXiv preprint*.

Liu, J., Y. Qian, Y. Yang, and Z. Yang (2022): "Can Artificial Intelligence Improve the Energy Efficiency of Manufacturing Companies? Evidence from China," *International Journal of Environmental Research and Public Health*, 19, 2091.

Mais Floresta (2023): "Pixlog lança tecnologia que revoluciona medição de madeira," `https://www.maisfloresta.com.br/pixlog-lanca-tecnologia-que-revoluciona-medicao-de-madeira/`, accessed: 2026-01-23.

Mamykina, L., B. Manoim, M. Mittal, G. Hripcsak, and B. Hartmann (2011): "Design Lessons from the Fastest Q&A Site in the West," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, CHI '11, 2857–2866.

Marr, B. and M. Ward (2019): *Artificial Intelligence in Practice: How 50 Successful Companies Used AI and Machine Learning to Solve Problems*, Hoboken, NJ: John Wiley & Sons.

Martinez, J. (2021): "Putty-Clay Automation," *CEPR Discussion*.

Maslej, N., L. Fattorini, R. Perrault, Y. Gil, V. Parli, N. Kariuki, E. Capstick, A. Reuel, E. Brynjolfsson, J. Etchemendy, K. Ligett, T. Lyons, J. Manyika, J. C. Niebles, Y. Shoham, R. Wald, T. Walsh, A. Hamrah, L. Santarlasci, J. B. Lotufo, A. Rome, A. Shi, and S. Oak (2025): "The AI Index 2025 Annual Report," Annual report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, retrieved 9 April 2025 from `https://hai-production.s3.amazonaws.com/files/hai_ai_index_report_2025.pdf`.

McElheran, K., J. F. Li, E. Brynjolfsson, Z. Kroff, E. Dinlersoz, L. S. Foster, and N. Zolas (2023): "AI Adoption in America: Who, What, and Where," Working Paper 31788, National Bureau of Economic Research.

McElheran, K., M.-J. Yang, Z. Kroff, and E. Brynjolfsson (2025): "The Rise of Industrial AI in America: Microfoundations of the Productivity J-curve(s)," Working Papers 25-27, Center for Economic Studies, U.S. Census Bureau.

McInnes, L., J. Healy, N. Saul, and L. Grossberger (2018): "UMAP: Uniform Manifold Approximation and Projection," *Journal of Open Source Software*, 3, 861.

Meyerovich, L. A. and A. S. Rabkin (2013): "Empirical analysis of programming language adoption," in *Proceedings of the 2013 ACM SIGPLAN International Conference on Object Oriented Programming Systems Languages & Applications*, New York, NY, USA: Association for Computing Machinery, OOPSLA '13, 1–18.

Miroudot, S., J. Sauvage, and B. Shepherd (2010): "Measuring the Cost of International Trade in Services," MPRA Paper 27655, University Library of Munich, Germany.

Murtaza, A., A. Saher, M. Zafar, S. Moosavi, M. Aftab, and F. Sanfilippo (2024): "Paradigm shift for predictive maintenance and condition monitoring from Industry 4.0 to Industry 5.0: A systematic review, challenges and case study," *Results in Engineering*, 102935.

Nanz, S. and C. A. Furia (2015): "A Comparative Study of Programming Languages in Rosetta Code," in *2015 IEEE/ACM 37th IEEE International Conference on Software Engineering*, IEEE, 778–788.

Navot, G. (2024): "The 8 Hidden Costs of Legacy Mainframe Codes in the Finance and Banking Sector," `https://overcast.blog/the-8-hidden-costs-of-legacy-mainframe-codes-in-the-finance-and-banking-sector-51e4e1ee62e3`, accessed: 16 Feb 2025.

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

Ovide, S. (2022): "The Long Life of COBOL: Why Banks and Governments Still Rely on It," https://www.nytimes.com/2022/07/06/technology/cobol-jobs.html, accessed: 16 Feb 2025.

Patel, H. and H. Ghelani (2024): "The Role of Artificial Intelligence (AI) in Quality Control of Industrial Products," ResearchGate, accessed: 2025-05-07.

Peng, S., E. Kalliamvakou, P. Cihon, and M. Demirer (2023): "The Impact of AI on Developer Productivity: Evidence from GitHub Copilot," .

Pixlog (2026): "Pixlog Manager," https://pixlog.com.br/, accessed: 2026-01-23.

Ponzanelli, L., G. Bavota, M. Di Penta, R. Oliveto, and M. Lanza (2014): "Mining StackOverflow to Turn the IDE into a Self-Confident Programming Prompter," in *Proceedings of the 11th Working Conference on Mining Software Repositories*, ACM, MSR 2014, 102–111.

Prytkova, E., F. Petit, D. Li, S. Chaturvedi, and T. Ciarli (2024): "The Employment Impact of Emerging Digital Technologies," CESifo Working Paper Series 10955, CESifo.

Rabah, S., J. Li, M. Liu, and Y. Lai (2010): "Comparative Studies of 10 Programming Languages within 10 Diverse Criteria – a Team 7 COMP6411-S10 Term Report," .

Reimers, N. and I. Gurevych (2019): "Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, Association for Computational Linguistics, 3982–3992.

——— (2020): "Making Monolingual Sentence Embeddings Multilingual using Knowledge Distillation," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, 4512–4525.

Ruckdeschel, M. (2025): "Just Read the Codebook! Make Use of Quality Codebooks in Zero-Shot Classification of Multilabel Frame Datasets," in *Proceedings of the 31st International Conference on Computational Linguistics*, ed. by O. Rambow, L. Wanner, M. Apidianaki, H. Al-Khalifa, B. D. Eugenio, and S. Schockaert, Abu Dhabi, UAE: Association for Computational Linguistics, 6317–6337.

Salton, G. and C. Buckley (1988): "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, 24, 513–523.

Schmitz, J. A. (2005): "What Determines Productivity? Lessons from the Dramatic Recovery of the U.S. and Canadian Iron Ore Industries Following Their Early 1980s Crisis," *Journal of Political Economy*, 113, 582–625.

Schwarz, C. (2019): "lsemantica: A command for text similarity based on latent semantic analysis," *The Stata Journal*, 19, 129–142.

Shi, W., X. Li, J. Zhang, and J. Sun (2023): "Design and Implementation of an AI-Based Quality Control System in Automotive Production," *Procedia Computer Science*, 228, 2322–2328.

Shinohara, E. and Y. Kawazoe (2026): "Efficient medical NER with limited data: Enhancing LLM performance through annotation guidelines," *International Journal of Medical Informatics*, 208, 106230.

Shojaeinasab, A. et al. (2022): "Intelligent Manufacturing Execution Systems: A Systematic Review," *Journal of Manufacturing Systems*, 62, 503–522.

Sullivan, M. (2025): "Elon Musk, Dogecoin, and the COBOL Programming Language," `https://www.fastcompany.com/91278597/elon-musk-doge-cobol-language`, accessed: 16 Feb 2025.

Tambad, S., R. Nandwani, and S. K. McIntosh (2020): "Analyzing programming languages by community characteristics on Github and StackOverflow," .

Tariq, U., A. Khan, S. Shah, M. Imran, et al. (2024): "A Novel Approach for Automated Quality Inspection Using Deep Learning-Based Computer Vision in Manufacturing Industry," *Electronics*, 13, 976.

Tbit (2024): "GroundEye – Rações," PDF brochure, produced by Tbit. PDF available on tbit.com.br (path includes /2024/08/).

TIOBE (2025): "TIOBE Programming Community Index," `https://www.tiobe.com/tiobe-index/`, accessed: 2025-07-18.

Torquato, P. (2022): "Importance of Registering a Software in Brazil," Accessed: 2024-11-14.

Treude, C. and M. P. Robillard (2016): "Augmenting API Documentation with Insights from Stack Overflow," in *Proceedings of the 38th International Conference on Software Engineering*, ACM, ICSE '16, 392–403.

Ucar, A., M. Karakose, and N. Kirimça (2024): "Artificial Intelligence for Predictive Maintenance Applications: Key Components, Trustworthiness, and Future Trends," *Applied Sciences*, 14, 898.

Vijayaraghavan, S., J. Song, T. Guan, S. Choi, and S. Kulkarni (2022): "Influence of Communication Among Shared Developers on the Productivity of Open Source Software Projects," .

Webb, M. (2020): "The Impact of Artificial Intelligence on the Labor Market," Retrieved from `https://www.michaelwebb.co/webb_ai.pdf`.

Wu, Y., S. Wang, C.-P. Bezemer, and K. Inoue (2019): "How Do Developers Utilize Source Code from Stack Overflow?" *Empirical Software Engineering*, 24, 637–673.

Xu, M., J. Liu, Y. Liu, F. X. Lin, Y. Liu, and X. Liu (2018): "A First Look at Deep Learning Apps on Smartphones," *arXiv preprint*.

Yang, F., Y. Liang, and X. Zhou (2023): "Learning Quality-Aware Representations for AI-Based Inspection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, 10506–10514.

A Theoretical Model

A.1 Equilibrium Definition

This appendix provides the formal details for the model in Section 2 of the main text. There is one final good, used as the numeraire. The wage and employment of skill group s in occupation o are $w_{o,s}$ and $L_{o,s}$.

A.2 Agents

The final-good producer chooses broad-group outputs to solve

$$\max_{\{Y_g\}_{g\in\mathcal{G}}} PY - \sum_{g\in\mathcal{G}} P_g Y_g,$$

where

$$Y = \left(\sum_{g\in\mathcal{G}} \varphi_g^{1/\chi} Y_g^{\frac{\chi-1}{\chi}}\right)^{\frac{\chi}{\chi-1}}, \qquad \varphi_g > 0, \qquad \sum_g \varphi_g = 1.$$

The broad-group producer chooses task quantities to solve

$$\max_{\{y(j)\}_{j\in J_g}} P_g Y_g - \int_{J_g} p(j)y(j)\,dj,$$

where

$$Y_g = \left(\int_{J_g} y(j)^{\frac{\psi-1}{\psi}}\,dj\right)^{\frac{\psi}{\psi-1}}.$$

Workers in occupation-skill cell (o, s) supply labor according to

$$L^S_{o,s} = \bar{L}_{o,s} w^{\eta}_{o,s}.$$

To supply $K_o(j)$, a developer solves

$$\min_{\{x_{o,l}(j)\}_{l\in\mathcal{L}}} \sum_{l\in\mathcal{L}} c_l x_{o,l}(j) \quad \text{s.t.} \quad K_o(j) \le \prod_{l\in\mathcal{L}} x_{o,l}(j)^{\alpha_{j,l}}.$$

Task $j \in J_o$ may be produced by either skill group. Conditional on skill s, task output is

$$y_{o,s}(j) = A_{o,s}(j)\left[\ell_{o,s}(j)^{\frac{\sigma_{g,s}-1}{\sigma_{g,s}}} + K_{o,s}(j)^{\frac{\sigma_{g,s}-1}{\sigma_{g,s}}}\right]^{\frac{\sigma_{g,s}}{\sigma_{g,s}-1}}.$$

For a required output level $y(j)$, the task producer solves

$$\min_{\ell_{o,s}(j),K_{o,s}(j)\ge 0} w_{o,s}\ell_{o,s}(j) + q_o(j)K_{o,s}(j)$$

subject to $y(j) \le y_{o,s}(j)$. The producer assigns the task to the least-cost skill group.

A.3 Equilibrium

Definition A.1 (Competitive equilibrium)**.** A competitive equilibrium is given by prices

$$\{P, \{P_g\}_{g\in\mathcal{G}}, \{p(j)\}_j, \{q_o(j)\}_{o,j}, \{w_{o,s}\}_{o,s}\},$$

allocations

$$\{Y, \{Y_g\}_g, \{y(j)\}_j, \{\ell_{o,s}(j), K_{o,s}(j)\}_{o,s,j}, \{K_o(j)\}_{o,j}, \{x_{o,l}(j)\}_{o,j,l}, \{L_{o,s}\}_{o,s}\},$$

and a task assignment $s^*(o,j)$ such that:

1. the final-good producer solves its problem, taking $\{P_g\}_g$ and P as given;
2. each broad-group producer solves its problem, taking $\{p(j)\}_{j\in J_g}$ and P_g as given;
3. labor supply satisfies
$$L^S_{o,s} = \bar{L}_{o,s} w^{\eta}_{o,s} \qquad \text{for all } o, s;$$
4. AI developers solve their cost-minimization problems and price AI at unit cost:
$$q_o(j) = \prod_{l\in\mathcal{L}} \left(\frac{c_l}{\alpha_{j,l}}\right)^{\alpha_{j,l}} \qquad \text{for all } o, j;$$
5. task producers solve their cost-minimization problems, task prices equal minimized unit costs, and tasks are assigned to least-cost skill groups:
$$s^*(o,j) \in \arg\min_{s\in\{L,H\}} p_{o,s}(j), \qquad p(j) = \min_{s\in\{L,H\}} p_{o,s}(j),$$
where $p_{o,s}(j)$ is the minimized unit cost for task j with skill group s. Inputs of non-assigned skill groups are zero:
$$\ell_{o,s}(j) = K_{o,s}(j) = 0 \qquad \text{if } s \neq s^*(o,j);$$
6. labor markets clear:
$$L^S_{o,s} = \int_{J_o} \mathbf{1}\{s = s^*(o,j)\}\ell_{o,s}(j)\, dj \qquad \text{for all } o, s;$$

7. AI markets clear task by task:

$$K_o(j) = \sum_{s \in \{L,H\}} \mathbf{1}\{s = s^*(o,j)\} K_{o,s}(j) \qquad \text{for all } o, j;$$

8. the final-good market clears and the numeraire normalization is

$$P = 1.$$

A.4 Equilibrium Characterization

From the first order conditions of developers, the price of AI is

$$q_o(j) = \prod_{l \in \mathcal{L}} \left(\frac{c_l}{\alpha_{j,l}} \right)^{\alpha_{j,l}} = \kappa_j \prod_{l \in \mathcal{L}} c_l^{\alpha_{j,l}}, \qquad \kappa_j \equiv \prod_{l \in \mathcal{L}} \alpha_{j,l}^{-\alpha_{j,l}}.$$

From firm's maximization, the unit cost of producing task $j \in J_o$ with skill group s is

$$p_{o,s}(j) = \frac{1}{A_{o,s}(j)} \left[w_{o,s}^{1-\sigma_{g,s}} + q_o(j)^{1-\sigma_{g,s}} \right]^{\frac{1}{1-\sigma_{g,s}}}.$$

If task j is produced with skill group s, labor and AI demand are given by

$$\ell_{o,s}(j) = y(j) A_{o,s}(j)^{\sigma_{g,s}-1} p(j)^{\sigma_{g,s}} w_{o,s}^{-\sigma_{g,s}} \tag{9}$$

$$K_{o,s}(j) = y(j) A_{o,s}(j)^{\sigma_{g,s}-1} p(j)^{\sigma_{g,s}} q_o(j)^{-\sigma_{g,s}}. \tag{10}$$

The corresponding labor and AI cost shares are

$$s_{L,o,s}(j) = \frac{w_{o,s}^{1-\sigma_{g,s}}}{w_{o,s}^{1-\sigma_{g,s}} + q_o(j)^{1-\sigma_{g,s}}}, \qquad s_{M,o,s}(j) = \frac{q_o(j)^{1-\sigma_{g,s}}}{w_{o,s}^{1-\sigma_{g,s}} + q_o(j)^{1-\sigma_{g,s}}}. \tag{11}$$

The equilibrium wage is given by

$$\bar{L}_{o,s} w_{o,s}^{\eta} = \int_{J_o} \mathbf{1}\{s = s^*(o,j)\} y(j) A_{o,s}(j)^{\sigma_{g,s}-1} p(j)^{\sigma_{g,s}} w_{o,s}^{-\sigma_{g,s}} \, dj. \tag{12}$$

A.5 AI Price Shocks and Exposure

Assume an AI price shock as in Equation (2). Also assume that the price change is small enough not to change the allocation of skill groups across tasks and to fall equally across

skill groups. Let

$$L_o = \sum_s L_{o,s}, \qquad \lambda_o(j) = \frac{\sum_s \mathbf{1}\{s = s^*(o,j)\}\ell_{o,s}(j)}{L_o}.$$

Write $s_{M,o}(j)$ for the AI cost share of the skill group assigned to task j. Occupation-level AI shock exposure is

$$\textit{AI Shock Exposure}_o = \int_{J_o} s_{M,o}(j)\lambda_o(j)\tau_o(j)\,dj. \tag{13}$$

Equivalently,

$$\textit{AI Shock Exposure}_o = \sum_{l\in\mathcal{L}} \Omega_{o,l}\tau_l, \qquad \Omega_{o,l} = \int_{J_o} s_{M,o}(j)\lambda_o(j)\alpha_{j,l}\,dj. \tag{14}$$

Let $J_{o,s}$ be the tasks assigned to skill group s and occupation o. Define

$$L_{o,s} = \int_{J_{o,s}} \ell_{o,s}(j)\,dj, \qquad \lambda_{o,s}(j) = \frac{\ell_{o,s}(j)}{L_{o,s}}, \tag{15}$$

$$\bar{s}_{M,o,s} = \int_{J_{o,s}} \lambda_{o,s}(j)s_{M,o,s}(j)\,dj, \qquad \mathcal{E}^{AI}_{o,s} = \int_{J_{o,s}} \lambda_{o,s}(j)s_{M,o,s}(j)\tau_o(j)\,dj. \tag{16}$$

For tractability, I assume that the AI shock is balanced across skill groups:

$$\mathcal{E}^{AI}_{o,L} = \mathcal{E}^{AI}_{o,H} \equiv \textit{AI Shock Exposure}_o \equiv \mathcal{E}_o. \tag{17}$$

Under the assumption of a small price shock, the change in labor is given by

$$d\ln L_{o,s} = \frac{\eta}{\eta + \psi + (\sigma_{g,s} - \psi)\bar{s}_{M,o,s}} \left[\Xi_g + (\psi - \sigma_{g,s})\mathcal{E}^{AI}_{o,s}\right]. \tag{18}$$

where

$$\Xi_g \equiv d\ln Y_g + \psi\, d\ln P_g. \tag{19}$$

Under the small-AI-share approximation, (18) becomes

$$d\ln L_{o,s} \approx \frac{\eta}{\eta+\psi}\Xi_g + \frac{\eta(\psi - \sigma_{g,s})}{\eta+\psi}\mathcal{E}^{AI}_{o,s}. \tag{20}$$

A.6 Proofs

Let

$$\mathcal{E}_o \equiv \textit{AI Shock Exposure}_o, \qquad \pi_o \equiv \frac{L_{o,L}}{L_o}, \qquad 1-\pi_o \equiv \frac{L_{o,H}}{L_o}.$$

Restatement of Proposition 1: AI has an ambiguous effect on employment. Under the local small-shock assumptions,

$$d\ln L_o \approx \frac{\eta}{\eta+\psi}\Xi_g + \beta_o^L \mathcal{E}_o,$$

where

$$\beta_o^L = \frac{\eta}{\eta+\psi}\left[\psi - \bar{\sigma}_g\right], \qquad \bar{\sigma}_g \equiv \pi_o\sigma_{gL} + (1-\pi_o)\sigma_{gH}.$$

Proof. For each skill group,

$$d\ln L_{o,s} \approx \frac{\eta}{\eta+\psi}\Xi_g + \frac{\eta(\psi-\sigma_{g,s})}{\eta+\psi}\mathcal{E}_o.$$

Since $L_o = L_{o,L} + L_{o,H}$,

$$d\ln L_o = \pi_o d\ln L_{o,L} + (1-\pi_o) d\ln L_{o,H}.$$

Substitution gives

$$d\ln L_o \approx \frac{\eta}{\eta+\psi}\Xi_g + \frac{\eta}{\eta+\psi}\left[\pi_o(\psi-\sigma_{gL}) + (1-\pi_o)(\psi-\sigma_{gH})\right]\mathcal{E}_o.$$

Using $\pi_o + (1-\pi_o) = 1$,

$$d\ln L_o \approx \frac{\eta}{\eta+\psi}\Xi_g + \frac{\eta}{\eta+\psi}\left[\psi - (\pi_o\sigma_{gL} + (1-\pi_o)\sigma_{gH})\right]\mathcal{E}_o.$$

The coefficient on exposure is positive if $\psi > \bar{\sigma}_g$, negative if $\psi < \bar{\sigma}_g$, and zero if $\psi = \bar{\sigma}_g$. □

Restatement of Proposition 2: Average Treatment Effect Depends on Sample Composition. The linear projection of occupation employment on AI exposure can be approximated by

$$d\ln L_o \approx \frac{\eta}{\eta+\psi}\Xi_g + \beta^L \textit{ AI Shock Exposure}_o + \epsilon_o,$$

where

$$\beta^L = \sum_o \omega_o \beta_o^L,$$

with $\pi_o, \omega_o \in (0,1)$, $\sum_o \omega_o = 1$, and ϵ_o orthogonal to *AI Shock Exposure*$_o$.

Proof. Let

$$\widetilde{Y}_o \equiv d\ln L_o - \frac{\eta}{\eta+\psi}\Xi_g, \qquad X_o \equiv \mathcal{E}_o.$$

Proposition 1 gives

$$\widetilde{Y}_o = \beta_o^L X_o.$$

The simple projection coefficient of $\widetilde{Y}_o$ on X_o is

$$\beta^L = \frac{\sum_o X_o \widetilde{Y}_o}{\sum_o X_o^2} = \frac{\sum_o X_o^2 \beta_o^L}{\sum_o X_o^2}.$$

Thus

$$\beta^L = \sum_o \omega_o \beta_o^L, \qquad \omega_o = \frac{X_o^2}{\sum_{o'} X_{o'}^2}.$$

If there is positive exposure variation, the weights are nonnegative and sum to one. The residual

$$\epsilon_o = \widetilde{Y}_o - \beta^L X_o$$

is orthogonal to X_o by the definition of the projection. □

Restatement of Proposition 3: AI can change the composition of workers within occupations. Assume that the initial AI expenditure share is small. Then, the effect of AI on the share of low-skilled workers is given by:

$$d\textit{Shr. Low-Skilled}_o \approx \Theta_o \left(\sigma_{g,H} - \sigma_{g,L}\right) \textit{AI Shock Exposure}_o,$$

where $\Theta_o > 0$ is a set of parameters.

Proof. The low-skill share is $\pi_o = L_{o,L}/L_o$. A first-order change is

$$d\pi_o = \pi_o(1-\pi_o)\left(d\ln L_{o,L} - d\ln L_{o,H}\right).$$

Using the skill-specific employment responses,

$$d\ln L_{o,L} - d\ln L_{o,H} \approx \frac{\eta}{\eta+\psi}\left[(\psi - \sigma_{g,L}) - (\psi - \sigma_{g,H})\right]\mathcal{E}_o.$$

The ψ terms cancel, so

$$d\ln L_{o,L} - d\ln L_{o,H} \approx \frac{\eta}{\eta+\psi}\left(\sigma_{g,H} - \sigma_{g,L}\right)\mathcal{E}_o.$$

Substituting into the share identity gives

$$d\pi_o \approx \frac{\eta}{\eta+\psi}\pi_o(1-\pi_o)\left(\sigma_{g,H} - \sigma_{g,L}\right)\mathcal{E}_o.$$

The stated expression follows with

$$\Theta_o = \frac{\eta}{\eta+\psi}\pi_o(1-\pi_o).$$

Since $\eta > 0$, $\psi > 0$, and $\pi_o \in (0,1)$, $\Theta_o > 0$. □

B Data

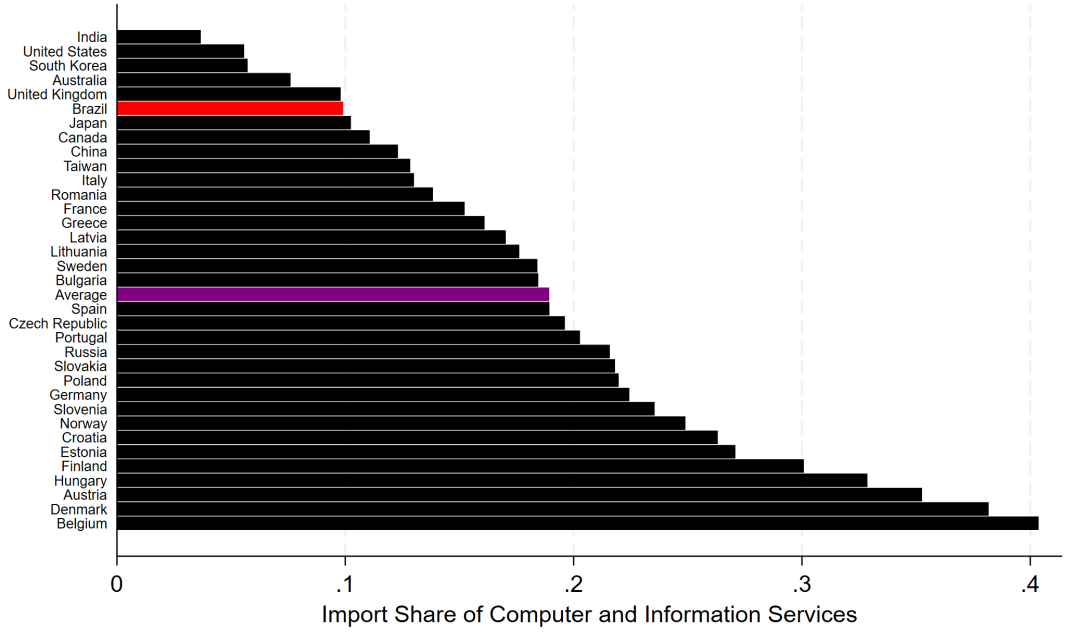
B.1 Most software used in Brazil is developed locally and therefore registered with the NIIP

Figure A1 reports the import share of computer services—a category that includes software publishing, programming, data processing, and all cross-border licensing fees for packaged software—using data from the WTO–OECD Balanced Trade in Services dataset combined with value-added data from the PREDICT database.[33] For Brazil, the import share of computer services is only 9.3%, compared to an OECD average of roughly 19%. Because imported software is typically paid for through licensing fees and does not appear in Brazil's copyright registry, the low import ratio suggests that the vast majority of software used by Brazilian firms is developed locally. As a result, the NIIP database, which records nearly the entire universe of domestic software, captures most of the software actually used in Brazil.

[33]Due to data restrictions, I cannot remove data processing and data center services from computer services. However, according to Junqueira Botelho et al. (2005), domestic software firms account for 98% of the Brazilian software market.

Figure A1: **Import Share of Computer Services Across the Globe in 2015**

In Brazil, Most Software is Locally Developed



Notes: This figure shows the import share of computer services across countries in 2015. The import share is calculated by dividing imports of computer services, which includes license fees paid to other countries, by the value added in computer services. Data on imports of computer services comes from the WTO-OECD Balanced Trade in Services dataset. Data on value added of computer services comes from the PREDICT Dataset.

One potential caveat is that foreign software vendors might operate through Brazilian subsidiaries. In such cases, inter-company transfers are recorded as royalties rather than licensing fees, and the parent firm might bypass Brazil's requirement to register source code locally. To assess the relevance of this channel, Table A1 reports the share of multinational enterprises (MNEs) in Brazil's IT sector: fewer than 1 percent of firms are foreign-controlled. This negligible presence suggests that nearly all software used in the country is produced by domestically owned firms and thus captured by the NIIP registry.

Table A1: **Share of Multinational Firms in IT Sectors**

	Share of Multinationals	*Employment-Weighted Share*
Information Technology Services	0.020%	0.025%
Information Service Provision Activities	0.038%	0.078%
Total	0.025%	0.035%

Notes: This table reports the share of foreign companies operating in the Brazilian IT sector in 2016. The sector *Information Technology Services* correspond to CNAE 2.0 code 62 and *Information Service Provision Activities* to CNAE code 63.

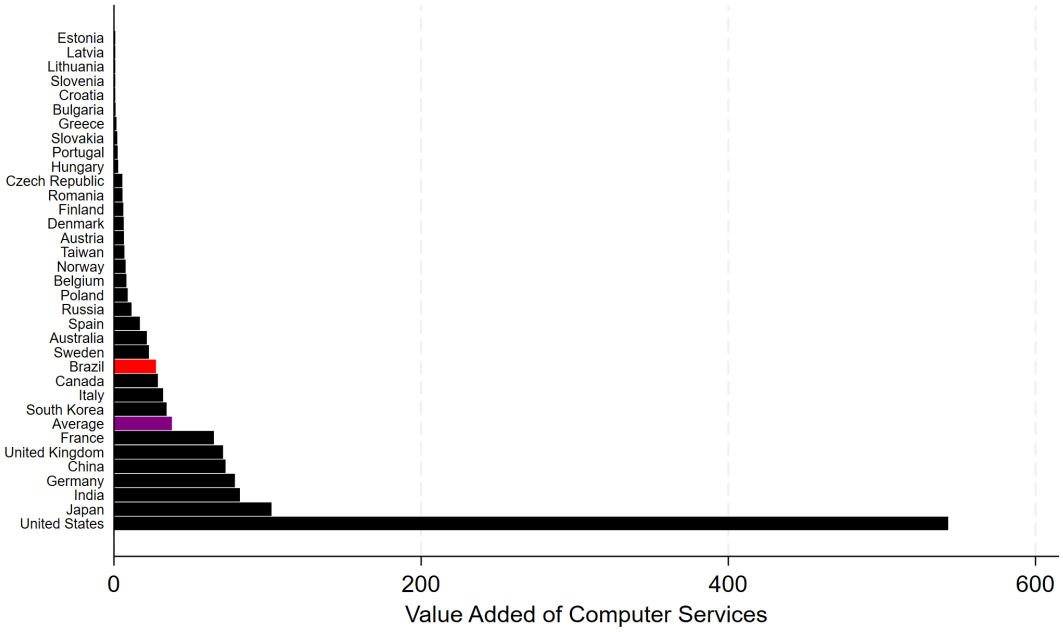
Three mutually reinforcing forces keep Brazil's import share of IT services exceptionally low. First, digital services tend to exhibit a strong home bias across countries. Alaveras and Martens (2015) shows that even in the United States, only the largest firms export online services, and that the preference for domestic providers is stronger for digital services than for physical goods. Trade costs in services are typically two to three times higher than in goods (Miroudot et al. 2010), partly due to cultural and regulatory frictions (Harms and Shuvalova 2020).

Second, Brazil reinforces this global home bias through explicit regulatory barriers. Local infrastructure mandates, data-localization requirements, and domestic-content rules in public procurement raise the effective trade cost of imported IT services to the equivalent of an 87 percent tariff—compared to an average of just 7 percent for goods trade (Ezell et al. 2013, Benz and Jaax 2022).

Third, Brazil has long fostered a large domestic software industry. Figure A2 shows that Brazil ranks as the 11th largest IT sector globally, ahead of developed countries such as Sweden, Australia, and Spain. As early as 2001, it ranked seventh worldwide, comparable to China and India (Junqueira Botelho et al. 2005). Large-scale public and private initiatives—such as social-security automation, electronic voting, and nationwide banking systems—generated sustained local demand (Arora and Gambardella 2005). As a result, 98 percent of software services in 2001 were supplied by Brazilian companies (Junqueira Botelho et al. 2005).

Taken together, the global difficulty of trading services, Brazil's unusually high regulatory barriers, and its mature domestic software industry explain why imported software plays only a minor role. Therefore, the NIIP registry provides near-census coverage of the code developed in Brazil.

Figure A2: **Value Added of Computer Services Across the Globe in 2015**



Notes: This figure shows value added of computer services across the globe in 2015 using data from the PREDICT Dataset in billions of 2015 dollars.

B.1.1 Survey with Industry Specialists

In this section I describe the results of a survey with industry specialists who have expert knowledge of the NIIP registration process. The survey delivers three main findings. First, software registration in the NIIP is widely recognized among both intellectual property lawyers and IT professionals, with the vast majority having direct experience with the registration process. Second, respondents confirm that most commercial software developed in Brazil is registered, with firms primarily motivated by intellectual property protection. Third, to the extent that selection exists, it is driven by small firms producing consumer-facing applications such as video games and mobile apps, categories that are unlikely to affect the labor market outcomes studied in this paper. These findings support the validity of using NIIP data to measure AI exposure in Brazilian workers.

Survey design. I conducted the survey with two groups that have detailed knowledge of the NIIP registration process: intellectual property lawyers and IT professionals. I identified potential respondents through an online search, constructing a database of 342 intellectual property law offices and 276 software development companies. I contacted these firms and of-

fices through email or LinkedIn between September and December of 2025. Each respondent was compensated for their participation.

In total, 68 people completed the survey: 41 intellectual property lawyers and 27 IT professionals. This gives a response rate of 11%. This small response rate comes from the fact that I contacted firms through publicly available email addresses, which often serve only customer inquiries rather than research requests. Individual lawyers, who typically manage their own correspondence, showed higher response rates than firms.

Survey respondents are experienced in their field. Table A2 presents summary statistics for the sample of respondents. Respondents average more than 10 years of experience in either IT or intellectual property law, with 86% holding at least a bachelor's degree. Among IT professionals, 81% work full-time in the IT sector, and 62% identify as programmers. Therefore, this is an experienced sample of respondents that are well-suited to assess the prevalence and patterns of software registration in Brazil.

Table A2: **Survey Summary Statistics**

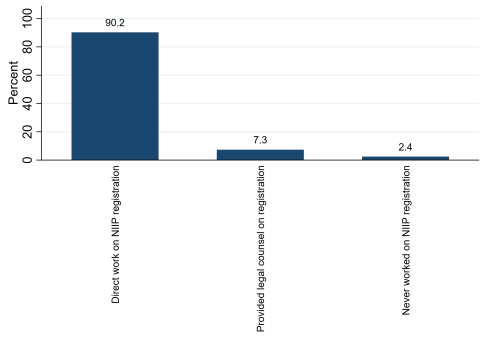
	Programmers	Lawyers	All
Average age	36.2	43.8	40.7
Average experience	12.9	18.9	16.6
Bachelor degree or more (%)	70.4	97.6	86.8

NIIP registration is recognized among specialists. The NIIP software registration process is widely known among industry specialists. Among IT professionals, 96% report being familiar with the registry. Among intellectual property lawyers, 90% have actively participated in at least one software registration, with the average lawyer having registered 37 distinct software programs. Among IT professionals, 77% report either having participated in a registration themselves or working at a firm that has registered software. This high level of familiarity indicates that the NIIP registry is a salient feature of Brazil's software industry, consistent with the institutional background described in Section 3.1.

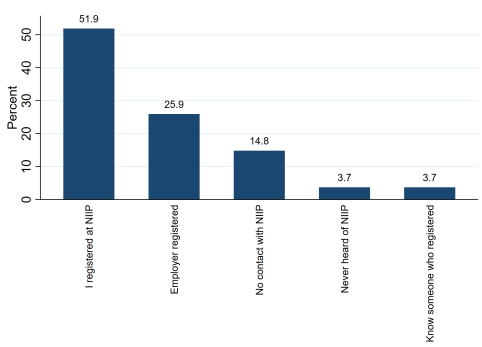
Figure A4a shows IT professionals' responses to the question: "How many commercial software programs do you believe have at least one version registered in the NIIP?" Seventy percent of respondents believe that firms register all or most of their commercial software,

Figure A3: **Experience of Respondents with a Software Registration**

(a) Intellectual Property Lawyer



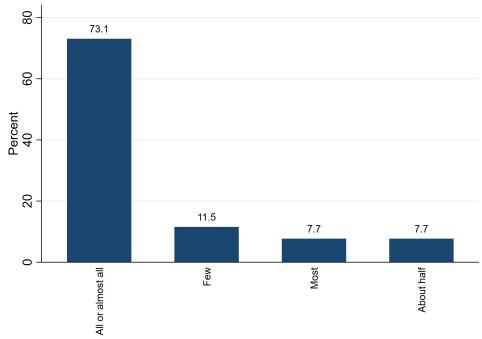
(b) IT Professional



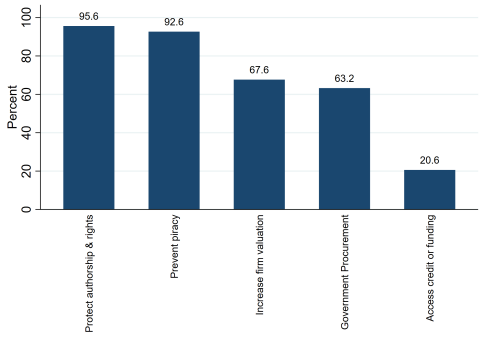
though 11% believe that only a few software programs are registered. This distribution suggests that while registration is the norm, some selection may exist.

Figure A4b displays the percentage of respondents who agree with various motivations for registering software in the NIIP. The dominant reasons are protecting intellectual property rights and preventing piracy, consistent with the main objectives of the NIIP.

(a) How many commercial software do you believe have at least one version registered in the NIIP?



(b) Why are firms registering software?



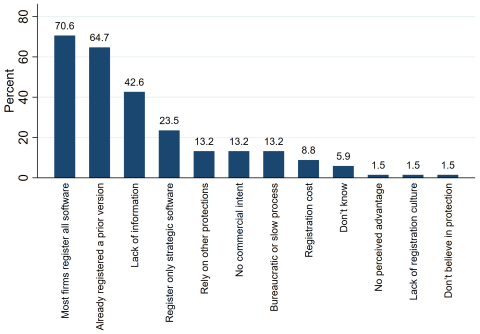
Selection is driven by small firms producing video games or cellphone apps. To understand which types of software might be underrepresented in the NIIP registry, I asked respondents about non-registration patterns. Figure A5a shows the reasons respondents cite

for why some software remains unregistered. The two most common explanations are that a previous version of the software has already been registered or that the firm is unaware of the NIIP. Notably, only 4% of IT professionals in the survey report being unfamiliar with the registry, suggesting that, although the lack of awareness explains why firms do not register, it is quantitatively a small factor.

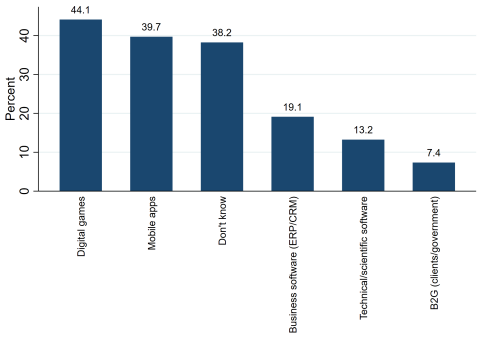
Figure A5b identifies which types of software are least likely to be registered. Respondents most frequently cite video games and mobile apps. Follow-up interviews with respondents reveal that developers of these consumer-facing applications perceive limited benefits from NIIP registration because distribution platforms (such as app stores) already provide intellectual property protections. Moreover, 83% of respondents agree that large companies are more likely to be aware of the NIIP and register their software than small firms.

Taken together, these findings indicate that any possible selection in the NIIP registry would be against small firms producing consumer-oriented applications. These categories represent a small share of enterprise software and are unlikely to significantly affect the labor market outcomes examined in this paper. The software that is used by firms, such as enterprise resource planning systems, manufacturing execution systems, and business intelligence tools, is comprehensively covered in the NIIP data.

(a) Why do some firms not register their software?



(b) What types of software do you think are less likely to be registered?



B.1.2 List of Broad Technical Classes and Application Domains

Table A3: Broad Technical Classes and Description

Code	Title	# Subcategories	Description
AP	Integrated Financial Suite	5	An Integrated Financial Suite is a comprehensive set of software tools designed to manage and streamline an organization's financial operations. It typically includes applications for planning budgets and projects, maintaining accounting records, performing audits, and overseeing internal controls. These tools work together to ensure financial accuracy, regulatory compliance, and strategic decision-making, providing a unified platform for managing the financial health and accountability of a business. Subcategories: Accounting, Application, Auditing, Control, and Planning.
AT	Automation Systems	8	Automation Systems are software-driven solutions that perform tasks with minimal human intervention, improving efficiency, consistency, and speed across various domains. This category spans office automation for routine administrative tasks, commercial and banking automation for streamlining business operations, and industrial automation for controlling manufacturing and production processes. It also includes specialized applications like automotive electronics, which manage vehicle systems. Subcategories: Automotive Electronics, Automation, Banking Automation, Commercial Automation, Industrial Automation, Manufacturing Automation, Office Automation, and Process Control.
AV	Resource Accounting	2	Resource Accounting tools monitor and evaluate how effectively software and hardware resources are utilized. This includes performance evaluation tools that test and measure system speed, responsiveness, and stability, as well as resource accounting tools that track the consumption of computing, energy, or other operational resources. These tools help identify inefficiencies, optimize performance, and support informed decision-making for system improvements and cost control. Subcategories: Performance Evaluation and Resource Accounting.
CD	Network Management	6	Network Management software enables the setup, monitoring, and coordination of computer networks and connected devices. It includes tools for managing local area networks (LANs), handling data communication between systems, and ensuring peripheral devices connect smoothly across the network. Terminal emulators allow users to access and control remote machines, while broader network management tools help maintain performance, security, and reliability. These systems are essential for keeping digital communication running smoothly, both within organizations and across distributed environments. Subcategories: Data Communication, Local Network, Network Management, Peripheral Management, Teleprocessing Monitor, and Terminal Emulators.
CT	Telecommunication Terminals	4	Telecommunication Terminals refer to systems and software that manage end-point communication devices and their supporting infrastructure. This includes telephone and telegraph switching systems that route calls and messages, as well as terminal operating systems that coordinate logistics and operations in terminal environments. Tools for function implementation and operations/maintenance management enhance performance and ensure system reliability. Together, these components optimize how terminals connect, operate, and are maintained within broader telecommunications networks. Subcategories: Function Implementation, Operations/Maintenance Management, Telephone/Telegraph Switching, and Terminal Operating System.
DS	Development Support Tools	8	Development Tools (Dev Tools) are essential software components that help programmers create, translate, and execute code efficiently. This category includes programming languages—both procedural and non-procedural—as well as the tools that transform source code into executable programs. Compilers, assemblers, and interpreters convert human-readable code into machine instructions, while pre-compilers, cross-compilers, and preprocessors adapt and optimize code for specific environments or hardware. Together, these tools form the technical foundation for software development, enabling developers to build, test, and maintain applications across platforms and architectures. Subcategories: Application Generator, Applications Development System, Computer Aided Software Engineering, Documentation Support, Programming Support, Routine Library, Systems Converter, and Systems Development Support Tool.
ET	Entertainment Software	4	Entertainment Software encompasses programs designed to engage users through interactive, visual, or immersive experiences. This includes video games, graphic art tools for creative expression, and virtual reality applications that simulate real or imagined environments. These tools are used for leisure, artistic development, and media production, blending technology with creativity to deliver engaging and often interactive content across platforms. Subcategories: Entertainment, Graphic Art, Video Game, and Virtual Reality.
FA	Office Applications	4	Office Applications are software tools designed to enhance productivity in everyday business and administrative tasks. This category includes word processors for document creation, spreadsheets for data analysis, and graphing tools for visualizing information. It also encompasses support tools like help desk software that assist in managing tasks and workflows. Together, these applications streamline communication, data handling, and problem-solving across a wide range of office environments. Subcategories: Electronic Spreadsheet, Graph Generator, Support Tool, and Word Processor.

Continued on next page

Code	Title	# Subcategories	Description
GI	Data Management	8	Data Management software supports the organization, storage, access, and integrity of digital information across its lifecycle. It includes tools for collecting and validating data, maintaining files, generating reports, and managing structured databases. Features like data dictionaries help define and standardize data use, while recovery tools ensure information can be restored after loss or corruption. The category also encompasses personal information managers and prototyping tools that aid in planning and interacting with data systems. Together, these tools enable individuals and organizations to handle data efficiently, securely, and reliably. Subcategories: Data Dictionary, Data Entry and Validation, Data Recovery, Database Manager, File Maintenance, Personal Information Manager, Report Generator, and Software Prototyping.
IA	AI Applications	3	AI Applications use intelligent algorithms to perform tasks that typically require human reasoning, learning, or perception. This includes expert systems that simulate decision-making in specialized fields, and natural language processing (NLP) tools that enable computers to understand, interpret, and generate human language in text or speech. Broadly, AI applications are used in areas such as healthcare, finance, customer service, and automation. Subcategories: Artificial Intelligence, Expert System, and Natural Language Processing.
IT	Measurement Tools	4	Measurement Tools includes software and systems designed to collect, analyze, and interpret data from various instruments. This includes general instrumentation software that controls and monitors devices, test and measurement tools used in scientific and industrial settings, and specialized biomedical instrumentation for healthcare applications. Analytical instrumentation further supports detailed examination and processing of data to derive meaningful insights. Together, these tools enable accurate and reliable measurement critical to research, diagnostics, and quality control across many fields. Subcategories: Analytical Instrumentation, Biomedical Instrumentation, Instrumentation, and Test and Measurement Instrumentation.
LG	Programming Language and Compilers	9	Programming Language and Compilers are essential software components that help programmers create, translate, and execute code efficiently. This category includes programming languages—both procedural and non-procedural—as well as the tools that transform source code into executable programs. Compilers, assemblers, and interpreters convert human-readable code into machine instructions, while pre-compilers, cross-compilers, and preprocessors adapt and optimize code for specific environments or hardware. Together, these tools form the technical foundation for software development, enabling developers to build, test, and maintain applications across platforms and architectures. Subcategories: Assembler, Compiler, Cross Compiler, Interpreter, Non-Procedural Language, Pre-Compiler, Pre-Processor, Procedural Language, and Programming Language.
PD	Data Protection	5	Data Protection software safeguards digital information from unauthorized access, loss, or corruption. This includes tools for securing data through encryption (cryptography), maintaining accuracy and consistency (data integrity), and controlling who can access or modify information (access control). Password managers help users create and store strong credentials securely, while broader security tools protect sensitive data across systems. Together, these solutions ensure the confidentiality, integrity, and availability of information in both personal and organizational settings. Subcategories: Access Control, Cryptography, Data Integrity, Password Manager, and Security and Data Protection.
SM	Simulation	4	Simulation software replicates real-world processes, systems, or environments to aid in analysis, training, or design. This category includes modeling tools for testing scenarios virtually, operating system simulators for educational or development purposes, and general simulators like those used in driving or flight training. It also encompasses engineering design and simulation tools such as CAE and CAM, which help optimize product designs and manufacturing processes. By allowing safe, cost-effective experimentation, simulation tools support better decision-making, improved performance, and innovation across various fields. Subcategories: Engineering Design/Simulation, Operating System Simulator, Simulation and Modeling, and Simulator.
SO	System Management	9	System Management software oversees the core functions that enable a computer or network to operate efficiently and reliably. It includes tools that manage hardware resources, coordinate input/output operations, and ensure seamless communication between system components. This category spans operating systems, device managers, user interfaces, and process controllers — all working together to allocate resources, run applications, and facilitate user interaction. System Management also covers lower-level interfaces for disks and networks, as well as control mechanisms that govern system behavior and automation. Subcategories: Basic Disk Interface, Command Processor, Communication Interface, Device Management, Input/Output Interface, Network Controller, Operating System, Process Controller, and User Interface Management.
TC	Image Processing	4	Image Processing software analyzes, enhances, and interprets visual data for a wide range of technical and scientific applications. It includes tools for digital image manipulation, pattern recognition, and operations research to extract meaningful information from images. These programs are used in fields such as medical imaging, remote sensing, and quality control, enabling automated analysis, improved accuracy, and deeper insights from visual inputs. Subcategories: Image Processing, Operations Research, Pattern Recognition, and Technical Scientific Application.

Continued on next page

Code	Title	# Subcategories	Description
TI	Telecom Systems	4	Telecom Systems support the exchange of information over distances through electronic means. This category includes technologies for data transmission, terminal devices that connect users to networks, and packet switching methods that efficiently route data across digital networks. Encompassing both hardware and software components, telecom systems form the foundation of modern communication—enabling voice, video, and data services across local and global networks. Subcategories: Data Transmission, Packet Switching, Telecommunication, and Terminals.
UT	Data Utilities	6	Data Utilities are supportive software tools designed to manage, process, and optimize data handling across computing systems. This includes utilities for compressing files, converting media formats, and transferring data between devices or networks. Tools like classifiers and interleavers help organize or structure data efficiently, while spool controllers manage queued tasks such as print jobs. Together, these utilities enhance system performance, streamline data workflows, and support seamless data exchange and transformation. Subcategories: Classifier/Interleaver, Data Compression, File Transfer, Media Conversion, Spool Controller, and Utilities.

Note: This table presents the code, title, number of subcategories, and description of each broad technical class. The description of each broad category was created by the authors using information from the original classification provided by the NIIP, Instituto Nacional da Propriedade Industrial (INPI) (2015a), and an inspection of the software in each classification.

Table A4: Broad Application Domains and Description

Code	Title	# Subcategories	Description
AD	Administration	11	The Administration application domain includes software that supports the management of public and private organizations. These systems handle strategic planning, operational control, and institutional development. Subfields include public administration, business management, production planning, personnel and material management, asset oversight, marketing, and office operations. Applications range from tools for organizing workflows and evaluating performance to systems for managing inventory, recruitment, sales strategies, and administrative communication. Subcategories: Organizational Development, Administrative Function and Planning, Modern Administration, Public Administration, Business Administration, Production Management, Personnel Management, Material Management, Asset Management, Marketing, Office Administration.
AG	Agriculture	14	The Agriculture application domain includes software that supports the planning, production, and management of agricultural systems. It encompasses crop cultivation, livestock farming, rural development, and agricultural policy. Applications address areas such as soil management, irrigation, pest control, and farm infrastructure through agricultural engineering tools. The domain also covers agricultural economics, aquaculture, and forestry, providing support for decision-making in resource use and production efficiency. Additionally, it includes systems for managing the extraction of plant and animal products, promoting sustainable use of natural resources. Subcategories: Agriculture and Rural Development, Agricultural Sciences, Agricultural Property Management, Agricultural Economics, Agricultural Systems, Agricultural Engineering, Soil Science, Plant Pathology, Crop Production, Animal Production and Veterinary Science, Forest Sciences, Aquaculture, Plant Extraction, Animal Extraction.
AH	Human Geography	6	The Human Geography application domain includes software that analyzes how human societies interact with space and place. It supports the study of social structures, cultural practices, urban and rural development, and the organization of communities. Applications cover topics such as migration, population distribution, urbanization, and public policy related to housing and infrastructure. This domain also incorporates tools for demographic analysis and urban geography, enabling users to model spatial patterns of human activity and assess their economic, social, and environmental impacts. Subcategories: Human Settlements, Urban Centers, Territorial Organization, Human Settlement Policies, Population Dynamics, Supporting Disciplines.
AN	Anthropology	7	The Anthropology application domain includes software that supports the study of human societies, cultures, and their evolution. It encompasses tools for analyzing social organization, community development, cultural traditions, and belief systems across different population groups. This domain integrates physical and cultural anthropology with sociology and the study of religion, enabling the examination of identity, social behavior, and institutional change. Applications often assist in ethnographic research, demographic studies, and comparative social analysis, offering insights into both historical and contemporary patterns of human life. Subcategories: Society and Social Structure, Social Development, Social Groups, Culture, Religion, Anthropology, Sociology.

Continued on next page

Code	Title	# Subcategories	Description
BL	Biology	8	The Biology application domain includes software that supports the study of living organisms and biological systems at all levels of organization. It spans molecular to systemic processes, aiding research in areas such as gene expression, cellular structure, microbial behavior, organ function, and metabolic regulation. Key subfields include genetics, cytology, microbiology, anatomy, physiology, biochemistry, and biophysics. These applications are used for modeling biological functions, analyzing experimental data, and simulating biochemical and physiological systems, contributing to advancements in medicine, environmental science, and biotechnology. Subcategories: Biology, Genetics, Cytology, Microbiology, Anatomy, Physiology, Biochemistry, Biophysics.
BT	Botany	4	The Botany application domain includes software that supports the study of plant life, focusing on plant structure, function, classification, and ecological distribution. Applications range from tools for analyzing plant morphology and physiology to systems for studying vegetation patterns across biomes. The domain also includes software for economic botany, which evaluates the properties and uses of crops, fibers, medicinal plants, and other economically important species. Taxonomic tools support plant identification and classification, while phytogeographic systems help map and understand regional vegetation. These applications are essential for biodiversity research, agriculture, and sustainable natural resource management. Subcategories: Botany, Phytogeography, Economic Botany, Plant Taxonomy.
CC	Construction	10	The Construction application domain includes software that supports the design, planning, execution, and management of building projects. It covers residential, commercial, industrial, and large-scale public works, incorporating traditional and modern construction techniques such as masonry, concrete, prefabrication, and mechanized processes. Applications assist with structural design, project coordination, cost estimation, bidding, and contract compliance, especially in public infrastructure projects. Specialized systems address structural analysis, finishing works, ventilation and lighting, hydraulic structures, and soil mechanics. These tools ensure efficient, safe, and sustainable construction practices across all phases of the building lifecycle. Subcategories: Construction, Construction Processes, Construction Management, Public Works, Structural Systems, Building Components, Construction Techniques, Building Hygiene, Hydraulic Engineering, Soil and Earthworks.
CO	Communication	6	The Communication application domain includes software that facilitates the creation, analysis, and dissemination of knowledge, language, and culture. It integrates tools from philosophy, science, and linguistics to explore how meaning is constructed and shared. Applications support mass media and public discourse through journalism, advertising, public relations, and editorial design. The domain also includes artistic expression—such as photography, cinema, music, and literature—as well as historical research and cultural preservation. These systems enable both practical communication tasks and broader engagement with cultural and intellectual life. Subcategories: Philosophy, Science and Scientific Methodology, Linguistics, Communication and Media, Arts, History.
DI	Legal	3	The Legal application domain includes software designed to support the creation, interpretation, and enforcement of laws across various levels of governance. It covers legislative drafting, legal analysis, and the application of constitutional principles that structure political institutions and protect individual rights. This domain also encompasses a wide range of legal disciplines, including civil, labor, tax, environmental, commercial, and international law. Applications assist with case management, contract generation, legal research, and compliance monitoring. These tools serve both public institutions and private entities, promoting legal order, protecting rights, and enabling the resolution of disputes across societal sectors. Subcategories: Legislation and Legal Interpretation, Constitutional Law, Specialized Legal Disciplines.
EC	Economics	14	The Economics application domain includes software that analyzes how individuals, firms, and governments allocate resources and respond to incentives. It covers both theoretical modeling and empirical analysis across microeconomics, macroeconomics, and national accounting. Applications support the study of production, consumption, and pricing, as well as broader economic dynamics like inflation, growth, and business cycles. The domain also includes tools for regional development, trade policy, monetary systems, income distribution, and firm behavior. These systems help inform decision-making and policy across local, national, and global economic contexts. Subcategories: Economic Theory and Methodology, Microeconomic Analysis, Macroeconomic Theory, Economic Activity, National Accounts, Monetary Economics, Market Structures, Products, Economic Engineering and Dynamics, Regional and Urban Economics, Property and Land Structure, International Economics, Economic Policy, Business and Corporation Finances.
ED	Education	6	The Education application domain includes software that supports the delivery, administration, and innovation of learning across formal and informal settings. It covers all levels of education, from early childhood to postgraduate studies, as well as adult education and vocational training. Applications assist with curriculum development, instructional methods, educational management, and policy implementation. The domain also includes tools for distance learning, multimedia instruction, and community-based programs. These systems help tailor educational experiences to diverse learners, improve institutional effectiveness, and expand access to knowledge. Subcategories: Regular Education, Supplementary Education, Educational Institutions and Administration, Instructional Methods and Materials, Curriculum and Academic Structure, Educational Systems and Policy.

Continued on next page

Code	Title	# Subcategories	Description
EL	Ecology	5	The Ecology application domain includes software that analyzes relationships between organisms and their environments across terrestrial, aquatic, and atmospheric systems. It supports studies in ecosystem dynamics, biodiversity, and environmental change. Subfields include ecophysiology, which examines how environmental factors influence biological function; human ecology, which focuses on sustainable development and urban systems; and plant and animal ecology, which explores species interactions, habitats, and ecological balance. Ethology applications further extend this domain by modeling behavioral patterns such as migration and hibernation. These tools are essential for research, conservation, and environmental management. Subcategories: General Ecology, Ecophysiology, Human Ecology, Plant and Animal Ecology, Ethology.
EN	Energy	6	The Energy application domain includes software that supports the production, distribution, and management of energy resources. It covers energy policy analysis, consumption tracking, and economic evaluation, as well as the technical aspects of energy systems. Applications assist in managing fossil fuels, biomass, nuclear power, and renewable energy sources. The domain also includes tools for energy conversion, storage, and distribution, as well as specialized areas such as microelectronics and nuclear reactor technology. Subcategories: Energy, Energy Resources, Fuels, Energy Technology, Electronics Engineering, Nuclear Engineering.
FN	Finance	6	The Finance application domain includes software that manages and analyzes financial activities in both the public and private sectors. It supports government budgeting, taxation, and fiscal administration, as well as corporate finance operations such as credit, banking, and capital markets. Applications also handle budgeting tools, investment tracking, resource allocation, and financial planning. The domain encompasses financial administration, including interest rate management, debt oversight, and financial risk control, as well as accounting systems for performance evaluation and compliance. Subcategories: Public Finance, Private Finance, Financial Systems, Budgeting and Financial Instruments, Financial Management, Accounting.
FQ	Physics-Chemistry	18	The Physics-Chemistry application domain includes software that supports the study of the fundamental properties of matter and energy and their interactions. It spans key areas of physics such as mechanics, thermodynamics, electromagnetism, waves, radiation, and particle dynamics, along with central fields of chemistry including organic, inorganic, analytical, and physical chemistry. Applications also cover spectroscopy, metrology, surface and molecular physics, and physical-chemical processes. Subcategories: Particle Physics, Acoustics and Optics, Wave Theory, Metrology, Mechanics, Solid and Fluid Physics, Thermodynamics, Electronics, Magnetism and Electromagnetism, Surface and Dispersion Physics, Radiation, Spectroscopy, Molecular Physics, General Chemistry, Analytical and Polymer Chemistry, Physical Chemistry, Organic Chemistry, Inorganic Chemistry.
GC	Geography	10	The Geography application domain includes software that supports the study and representation of Earth's physical and human landscapes. It covers physical geography, which analyzes landforms, geomorphology, and natural processes, and human geography, which explores population patterns, economic systems, and political organization. The domain also includes regional and geographic orientation studies, as well as geospatial sciences such as geodesy, topography, and remote sensing. Specialized tools for photogrammetry, mapping, and cartographic methods facilitate accurate spatial analysis and visualization. Subcategories: Physical Geography, Human Geography, Regional Geography, Geographic Orientation, Geodesy, Topography, Photogrammetry, Mapping, Cartographic Methods, and Cartographic Plans.
GL	Geology	7	The Geology application domain includes software that supports the analysis of Earth's structure, composition, and geological processes. It covers surface dynamics like erosion and weathering, internal processes such as tectonics and volcanism, and structural geology including folds and faults. Applications extend to glaciology and marine geology, as well as historical geology fields like paleontology and stratigraphy. Economic geology tools assist in locating and analyzing mineral deposits, while geochemistry, geophysics, hydrogeology, and geotechnics provide insights into subsurface materials and conditions. Subcategories: Physical Geology, Glaciology, Geotectonics, Marine Geology, Historical Geology, Economic Geology, and Geochemistry/Geophysics/Geotechnics.
HB	Real Estate	2	The Real Estate application domain includes software that supports the analysis, planning, and management of housing systems and residential environments. It addresses housing markets, supply and demand dynamics, and public housing policies. Applications also classify and manage different housing types, including single- and multi-family units, student residences, elderly housing, and temporary or informal dwellings. Subcategories: Real Estate Policy and Markets, Real Estate Typology.
HD	Hydrology	4	The Hydrology application domain includes software that supports the analysis and monitoring of water in the environment. It covers the hydrological cycle, tracking the movement and distribution of water across rivers, lakes, and the atmosphere. Applications also support hydrography, which maps water bodies and flood zones, and hydrometry, which involves measuring rainfall, streamflow, evaporation, and sediment transport. The domain extends to oceanography, encompassing the physical, chemical, biological, and geological study of marine environments. Subcategories: Hydrology, Hydrography, Hydrometrics, Oceanography.

Continued on next page

Code	Title	# Subcategories	Description
IF	Information Management	10	The Information Management application domain includes software that organizes, processes, and distributes data and knowledge across various platforms and institutions. It supports the management of scientific, technical, and strategic information through systems for storage, retrieval, analysis, and dissemination. This domain encompasses libraries, archives, and documentation centers, as well as reprography tools like photocopying and microfilming. It also includes information systems and networks designed to optimize data flow, user access, and service delivery. Subcategories: Information, Information Documentation, Reprography, Information Materials, Library Management, Archival Science, Information Science, Information Services, User Information, Data Processing.
IN	Manufacturing	5	The Manufacturing application domain includes software that supports industrial production, transformation processes, and technological development. It covers a broad range of sectors, including metallurgy, electronics, chemicals, textiles, and food processing, as well as extractive industries like mining. Applications assist with industrial policy planning, production monitoring, and innovation management. The domain also integrates engineering tools for product design, testing, and manufacturing efficiency, alongside systems for managing technological research and the adoption of appropriate production technologies. Subcategories: Manufacturing, Industrial Technology, Industrial Engineering, Extractive Industry, Transformation Industry.
MA	Environment	4	The Environment application domain includes software that supports the protection, management, and improvement of natural and built environments. It addresses the development of environmental policies, conservation of renewable and non-renewable resources, and the monitoring and control of pollution—including air, water, soil, and noise pollution. Applications also focus on environmental quality, using data and engineering solutions to assess and improve living conditions. Subcategories: Environment, Natural Resources, Pollution, Environmental Quality.
ME	Meteorology	3	The Meteorology application domain includes software that supports the study and forecasting of atmospheric conditions. It encompasses atmospheric physics and dynamics, focusing on variables such as air pressure, wind, temperature, humidity, and solar radiation. Applications also support climatology, which analyzes long-term climate trends, seasonal patterns, and their influence on agriculture and ecosystems. Subcategories: Meteorology, Atmospheric Science, Climatology.
MT	Mathematics	6	The Mathematics application domain includes software that supports the development and application of mathematical theories and methods. It spans foundational areas such as logic, algebra, geometry, analysis, and calculus, which form the basis of formal reasoning and quantitative problem-solving. Applied mathematics extends these tools to real-world contexts through modeling, statistics, probability, and operations research. Subcategories: Logic, Algebra, Geometry, Analysis, Calculus, Applied Mathematics.
PD	Pedology	3	The Pedology application domain includes software that analyzes the formation, structure, and classification of soils in their natural settings. It addresses physical, chemical, mineralogical, and biological soil properties, supporting studies of soil horizons and development processes through morphopedological analysis. Applications help identify and map soil types, aiding in agricultural planning, land use management, and environmental conservation. Subcategories: Soil Science, Soil Formation, Soil Types.
PL	Public Policy	2	The Public Policy application domain includes software that supports the analysis and development of governmental policies and political systems. It draws on political science to model institutional structures, understand governance mechanisms, and evaluate public decision-making processes. Applications explore theories of power, state organization, political regimes, and policy implementation. Subcategories: Public Policy Science, Public Policy Systems.
PR	Social Security	3	The Social Security application domain includes software that supports the delivery and management of social welfare programs. It covers social security systems, including public and private retirement and pension schemes, as well as various forms of social assistance such as healthcare, housing, food aid, and rehabilitative services. These applications help implement policies, track benefits, and coordinate services aimed at reducing vulnerability and improving quality of life. They are essential for ensuring equitable access to support systems and strengthening the social safety net through both public and private initiatives. Subcategories: Social Security, Retirement Benefits, Social Assistance.
PS	Psychology	3	The Psychology application domain includes software that supports the study and application of mental processes and human behavior. It covers areas such as developmental, clinical, social, and educational psychology, as well as the analysis of sensory, cognitive, and emotional functions. Applications also explore behavioral theories, including behaviorism, existentialism, and reinforcement models. These tools are used in research, therapy, education, and mental health services to better understand psychological functioning and promote well-being, learning, and personal growth. Subcategories: Psychology, Human Behavior, Psychological Theories.
SD	Health	11	The Health application domain includes software that supports the delivery, management, and advancement of healthcare services. It covers public and mental health policy, healthcare administration, and disease prevention and treatment. Applications range from medical diagnostics and therapies to systems for hospital, ambulatory, and home-based care. The domain also includes general and specialized medicine, pharmacology, dentistry, and biomedical sciences. Subcategories: Public Health, Health Administration, Disease, Disabilities, Medical Assistance, Medical Diagnostics, General Medicine, Medical Specialties, Biomedical Engineering, Pharmacology, Dentistry.

Continued on next page

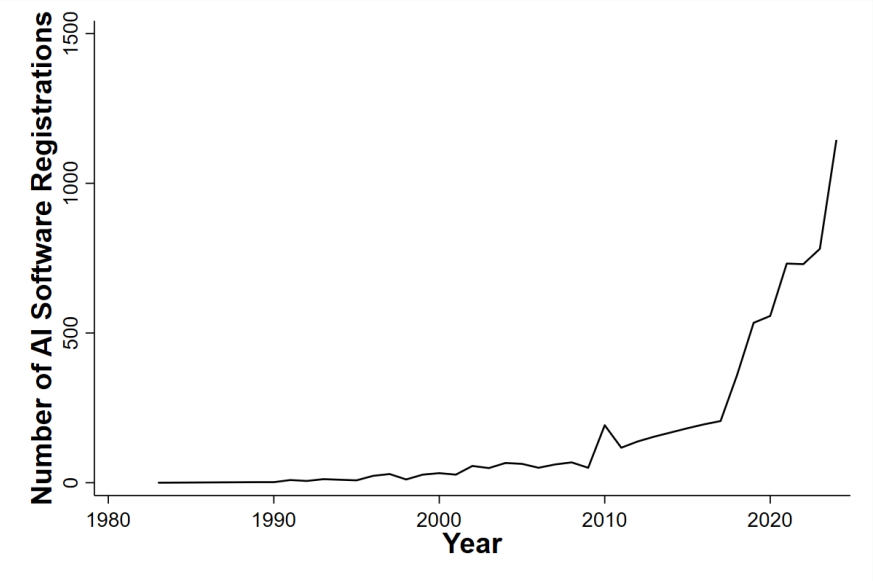
Code	Title	# Subcategories	Description
SM	Sanitation	5	The Sanitation application domain includes software that manages systems essential for public health and environmental quality. It encompasses sanitary engineering, basic sanitation services, and the treatment and disposal of various types of waste—solid, liquid, chemical, and industrial. Applications support public cleaning operations, including garbage collection and urban drainage, as well as water supply systems that handle the capture, treatment, and distribution of drinking water. The domain also covers sewage management, with tools for monitoring and treating both domestic and industrial wastewater. Subcategories: Sanitation Systems, Waste Management, Urban Cleaning, Water Supply, Wastewater and Sewage.
SV	Services	4	The Services application domain includes software that supports a wide array of public and private service activities essential to social and economic life. It encompasses utilities such as electricity, water, and telecommunications, as well as services related to maintenance, personal care, and public safety. Security applications manage insurance, risk assessment, and asset protection. Retail-focused tools handle trade operations, commercial logistics, and market policies, while tourism systems support travel planning, infrastructure management, and visitor services. Subcategories: Public and Private Services, Insurance, Commerce, Tourism.
TB	Human Resources	6	The Human Resources application domain includes software that manages the organization and development of labor and workforce systems. It covers different types of work—including manual, intellectual, rural, and technical—while addressing labor conditions, ergonomics, and employment policies. Applications support workforce planning, wage analysis, job classification, and labor market monitoring, as well as the structure of occupations through unions, councils, and professional associations. The domain also includes tools that promote worker well-being, such as leisure and recreation management. Subcategories: Organization of Work, Human Resources, Labor Market, Ergonomics, Occupational Structure and Professional Organizations, Leisure.
TC	Telecommunications	4	The Telecommunications application domain includes software that supports the planning, operation, and regulation of communication systems. It encompasses telecommunication policies and models, as well as technical systems such as telephony, television, radiocommunication, and data transmission. Engineering applications focus on the infrastructure for signal transmission and reception, including wired and wireless networks. The domain also covers the management of telecom services, network operations, and communication stations. Subcategories: Telecommunication Policy, Telecommunication Systems, Telecommunication Engineering, Network Services and Infrastructure.
TP	Logistics	5	The Logistics application domain includes software that coordinates the movement of goods and people across transport networks. It covers transport policy, infrastructure planning, and service operations across road, rail, air, and water modes. Applications support cargo and passenger management, vehicle routing, terminal operations, and multimodal integration. Logistics software also optimizes supply chains, enabling real-time tracking, route planning, and load management. Subcategories: Transport Policy and Planning, Transportation Networks and Infrastructure, Passenger and Freight Transport Services, Transport Engineering and Modal Systems, Modes of Transportation.
UB	Urban Planning	5	The Urban Planning application domain includes software that guides the development and organization of urban spaces. It supports land use planning, zoning, real estate regulation, and infrastructure design. Applications also address urban mobility, public services, and architectural projects across residential, commercial, and institutional buildings. Subcategories: Urban Planning and Design, Urban Land and Property, Urban Area Structure and Zoning, Urban Circulation and Infrastructure, Architecture.

Note: This table presents the code, title, number of subcategories, and description of each broad application domain. The description of each broad application domain was created by the authors using information from the original classification provided by the NIIP, Instituto Nacional da Propriedade Industrial (INPI) (2015b), and an inspection of the software in each classification.

B.2 Summary Statistics of the Software Dataset

This section presents summary statistics for the software dataset. Figure A6 displays the number of AI software registrations over time. Consistent with the trends discussed in Section 4, there is a sharp increase in AI software creation beginning around 2013.

Figure A6: **Number of New AI Software Registrations Over Time**

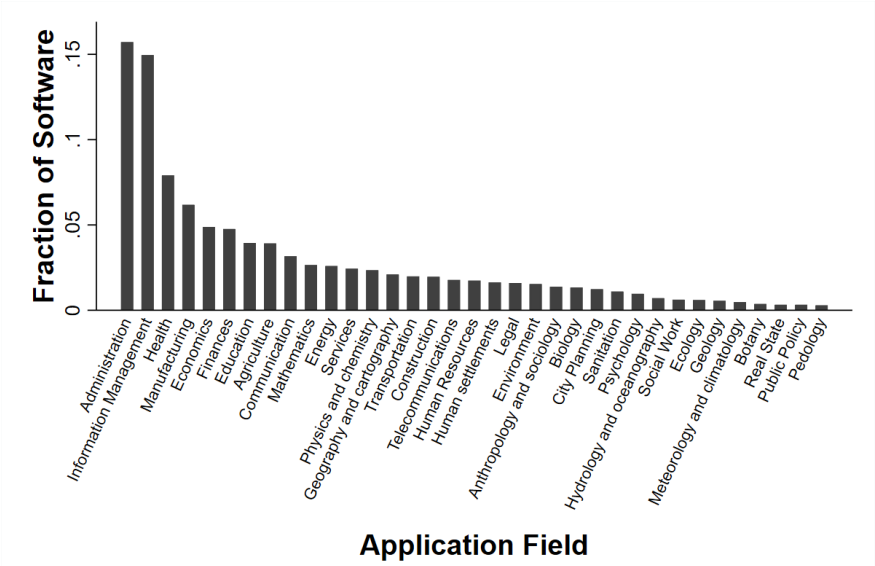


Notes: This figure plots the number of newly registered software classified as artificial intelligence (AI) in Brazil from 1987 to 2023.

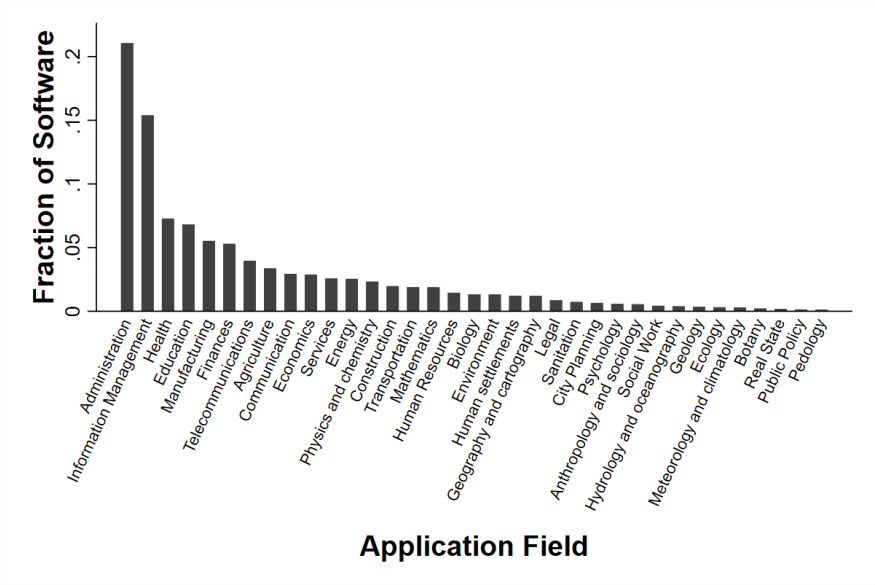
Figure A7 compares the distribution of application domains between AI and non-AI software. As shown in Figure A7b, non-AI software is more heavily concentrated in administrative and information management domains, whereas AI software is more broadly distributed across different areas.

Figure A7: **AI and Non-AI Application Domain**

(a) Fraction of Application Domain Among AI Software



(b) Fraction of Application Domain Among Non-AI Software

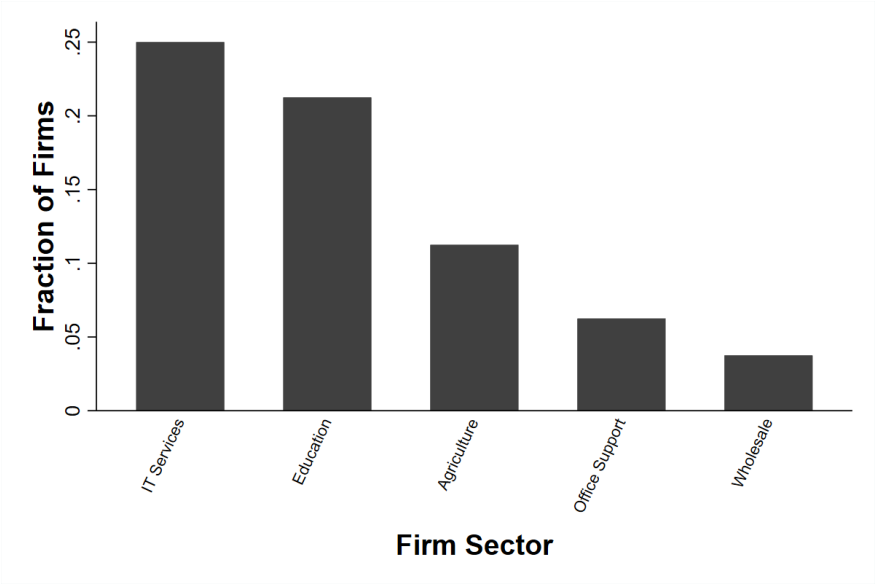


Notes: These figures present the distribution of AI software across different application domains, as defined by the NIIP. The application domain classification was developed by the NIIP to indicate the areas in which each software is intended to be used. Because a single software program may be assigned to multiple domains, each figure reports the fraction of each domain relative to the total number of domain classifications. Figure A7a plots the distribution of application domain among AI software while Figure A7b has the distribution of application domain among non-AI software.

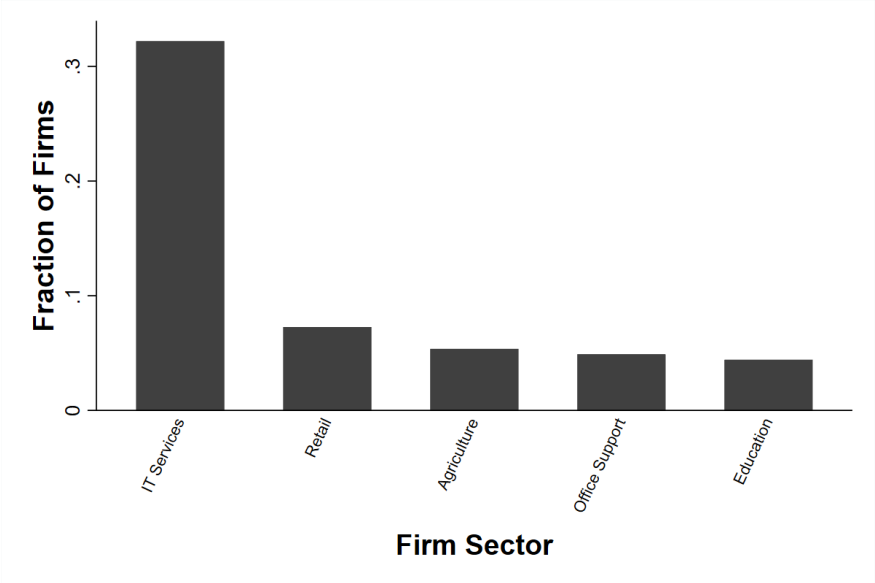
Figure A8 displays the distribution of AI and non-AI software across firm sectors. Among AI software, the most intensive sectors are Information Technology (IT) and Education, which includes universities and research institutions. In contrast, non-AI software is also concentrated in the IT sector but followed by the retail sector.

Figure A8: **Sectoral Distribution of Firms Creating AI and Non-AI Software**

(a) Sectoral Distribution of Firms Creating AI Software



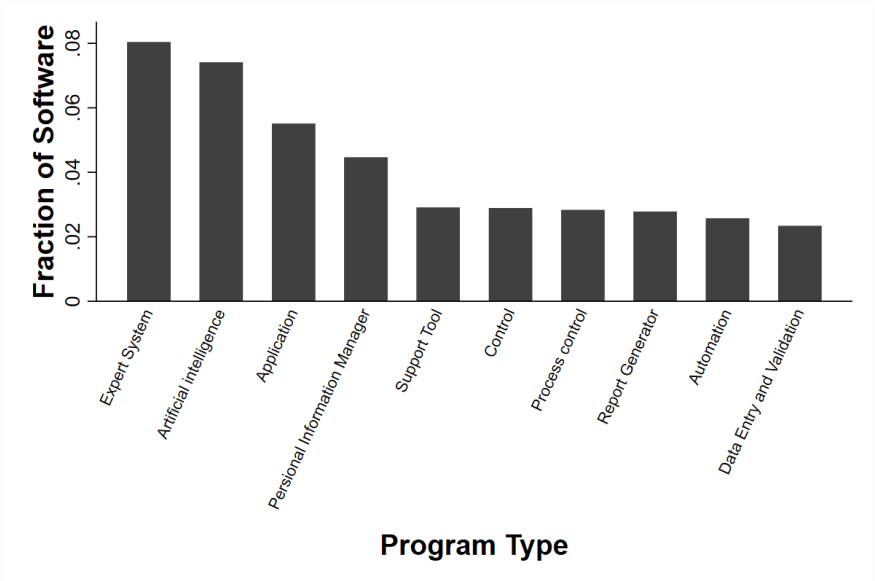
(b) Sectoral Distribution of Firms Creating non-AI Software



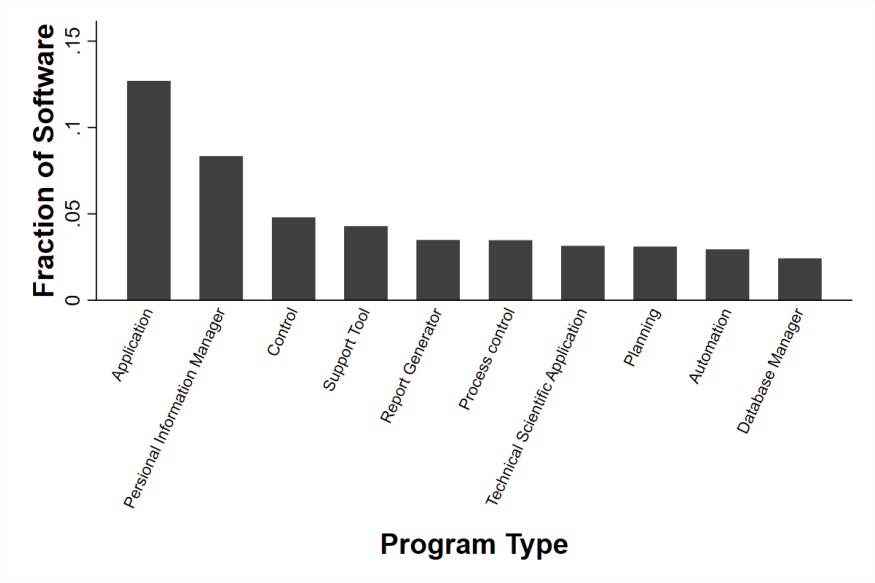
Notes: These figures show the distribution of firms registering AI and non-AI software across economic sectors.

Figure A9: **Fraction of Technical Class Among AI and Non-AI Software**

(a) Fraction of Technical Class Among AI Software



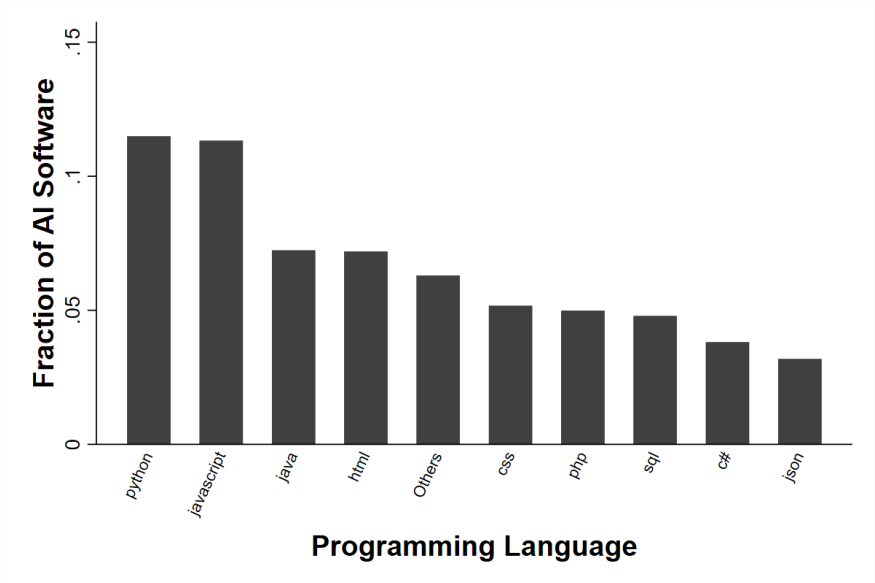
(b) Fraction of Technical Class Among non-AI Software



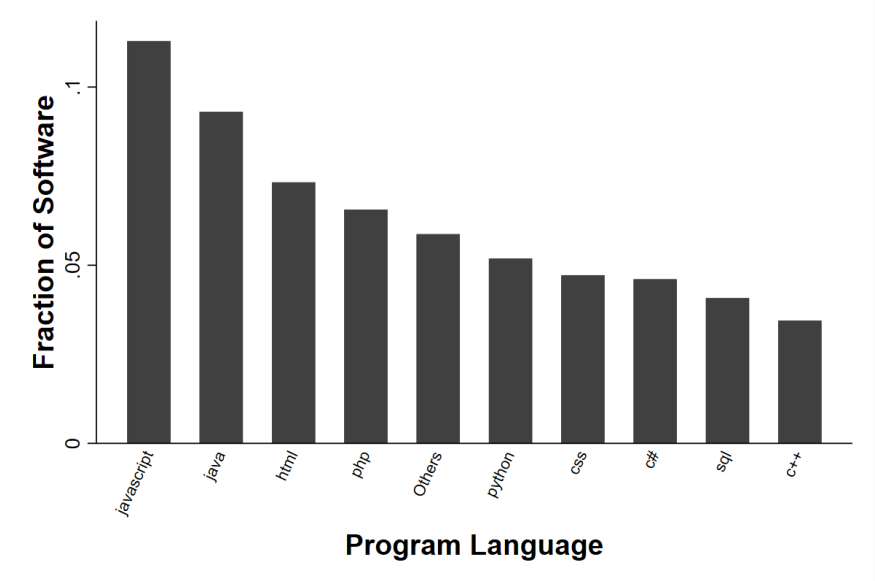
Notes: These figures show the distribution of technical class among AI and non-AI software.

Figure A10: **Fraction of Different Programming Languages Among AI and Non-AI Software**

(a) Fraction of Programming Language Among AI Software

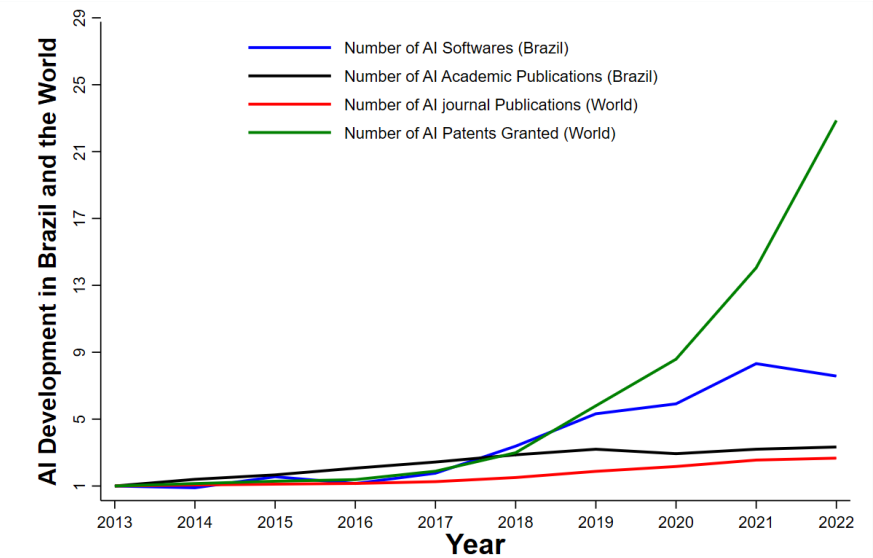


(b) Fraction of Programming Language Among non-AI Software



Notes: These figures show the distribution of programming languages among AI and non-AI software.

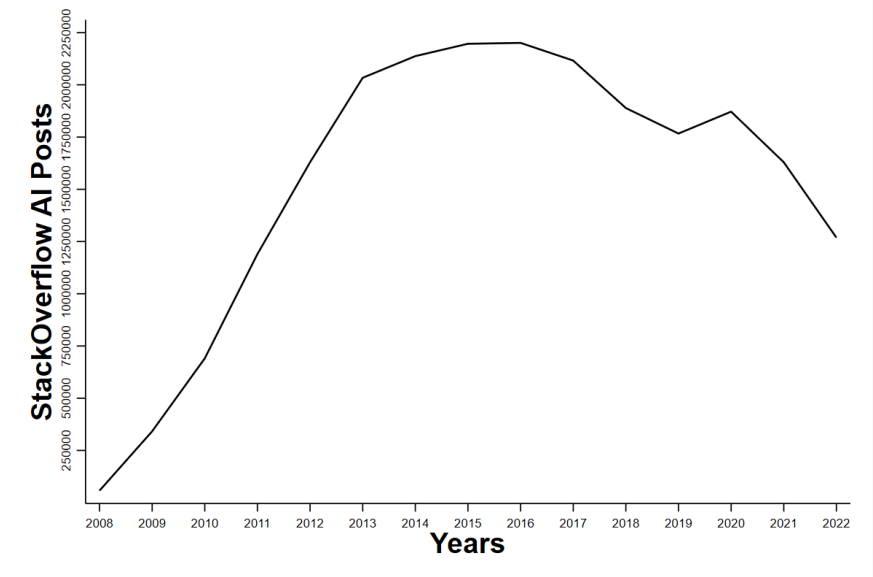
Figure A11: **AI Development in Brazil Follows Global Trend**



Notes: This figure plots the time series of four indicators of AI development: (i) the number of AI software programs registered in Brazil's NIIP database, (ii) the number of AI-related academic publications in Brazil, (iii) the number of AI-related journal articles globally, and (iv) the number of AI patents granted worldwide. Data for items (ii)–(iv) are from Maslej et al. (2025), while the software registration data are drawn from the NIIP registry.

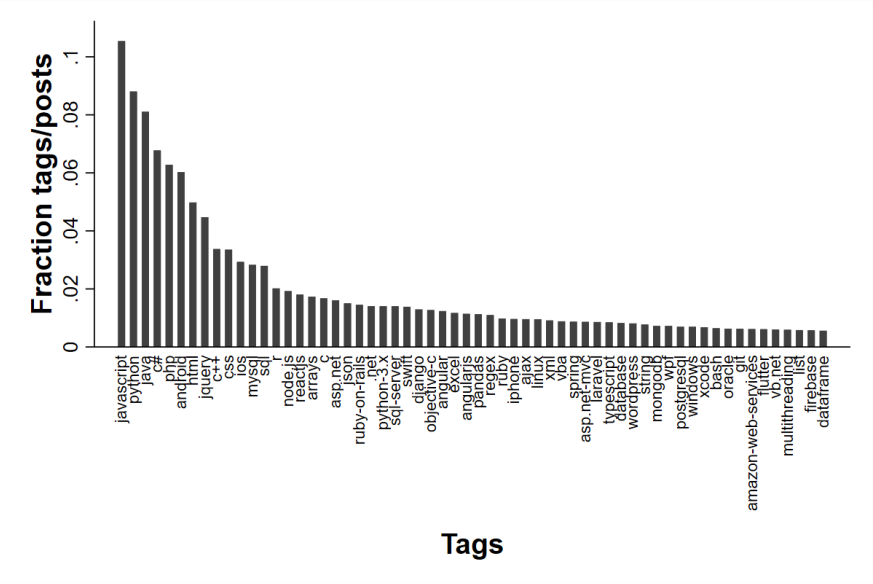
B.3 Statistics of Stack Overflow Posts

Figure A12: Number of stack overflow posts per year



Notes: This figure plots the number of new Stack Overflow posts over time.

Figure A13: Top 25 most popular tags



Notes: This figure plots the most common tags in StackOverflow and the share of posts with each one of them.

B.4 Specialization of Programming Languages

Table A5: **Main Programming Language for Each Technical Class**

Technical Class	Main Language	Share
Applications	javascript	.431
Planning	javascript	.494
Control	javascript	.534
Audit	javascript	.553
Accounting	javascript	.560
Automation	javascript	.440
Office Automation	javascript	.457
Commercial automation	javascript	.421
Banking Automation	javascript	.383
Industrial automation	javascript	.327
Process control	javascript	.541
Manufacturing Automation (CNC, Robotics)	javascript	.342
Automotive Electronics	c#	.419
Performance evaluation	javascript	.552

Continued on next page

Table A5 – continued from previous page

Technical Class	Main Language	Share
Resource Accounting	javascript	.587
Data communication	javascript	.517
Terminal Emulators	c#	.475
Teleprocessing Monitors	javascript	.442
Device and Peripheral Management	javascript	.447
Data Communication Network Manager	javascript	.450
Local network	javascript	.397
Telephone and Telegraph Switching	javascript	.500
Additional Functions Implementer	javascript	.475
Operation and Maintenance Manager	javascript	.510
Central Operation and Maintenance Terminal	javascript	.500
Systems Development Support Tools	javascript	.426
Application Generator	javascript	.469
Computer Aided Software Engineering	sql	.412
Applications Development System	javascript	.435
Routine Libraries ("Libraries")	javascript	.363
Programming Support	javascript	.320
Documentation Support	javascript	.441
Systems Converter	javascript	.509
Entertainment	javascript	.384
Animated Games ("arcade games")	c#	.548
Drawing Generators	c#	.450
Simulators for Leisure	c#	.457
Support Tool	javascript	.416
Word Processors	javascript	.500
Electronic Spreadsheets	javascript	.503
Graph Generators	javascript	.558
Information Manager	javascript	.463
Database Manager	javascript	.511
Screen Generator	javascript	.522
Report Generator	javascript	.536
Data dictionary	javascript	.578
Data Entry and Validation	javascript	.482
File Maintenance	javascript	.482
Data recovery	javascript	.401

Continued on next page

Table A5 – continued from previous page

Technical Class	Main Language	Share
Artificial intelligence	python	.328
Expert Systems	javascript	.321
Natural Language Processing Systems	python	.490
Instrumentation	c#	.321
Test and Measurement Instrumentation	javascript	.405
Biomedical Instrumentation	javascript	.321
Analytical Instrumentation	javascript	.532
Languages	javascript	.528
Compiler	javascript	.544
Assembler	java	.519
Pre-Compiler	javascript	.500
Cross Compiler	javascript	.519
Pre-Processor	java	.538
Interpreter	java	.421
Procedural Language	javascript	.500
Non-Procedural Language	javascript	.564
Security and Data Protection	javascript	.476
Password	javascript	.487
Cryptography	javascript	.519
Maintaining Data Integrity	javascript	.546
Access Control	javascript	.528
Simulation and Modeling	python	.297
Flight/Car/Submarine Simulator	c#	.485
Operating Environment Simulators	javascript	.417
CAE/CAD/CAM/CAL/CBT	c#	.419
Operational system	javascript	.375
Input and Output Interface	javascript	.471
Basic Disk Interface	python	.389
Communication Interface	javascript	.409
User Manager	javascript	.570
Device Administrator	javascript	.473
Process Controller	javascript	.457
Network Controller	sql	.365
Command Processor	java	.371
Technical-Scientific Applications	python	.365

Continued on next page

Table A5 – continued from previous page

Technical Class	Main Language	Share
Operational Research	python	.359
Pattern Recognition	python	.439
Image Processing	python	.427
Teleinformatics	javascript	.364
Terminals	sql	.450
Data Transmission	javascript	.479
Data Switching	javascript	.473
Utilities	javascript	.361
Data Compressor	javascript	.423
Storage Media Converter	javascript	.462
Classifier/Interleaver	python	.426
Spool Controller	c#	.444
File transference	javascript	.553

Figure A14: **Word Cloud from AI Software Titles**



Notes: This figure shows a word cloud from the titles of AI software.

C Empirics

C.1 Text Similarity

In this section, I describe in detail how I calculate the text similarity between software descriptions and occupational tasks.

Parsing. I transform documents into vectors using words as tokens, i.e., unigrams.

Selection. To avoid counting frequent and uninformative words, such as the Portuguese equivalent for "the" and "and", I drop terms that appear in more than 80% of documents.

Vectorization. Following the previous steps, I characterize each document with a vector of dummies for words it contains. Let $m \in \{1, ..., M\}$ index the set of words in the documents, $\mathcal{M}$. Let c_{km} be a dummy variable taking the value of 1 if the document k contains word m. Therefore, document k can be represented by vector c_k with entries c_{km}.

Normalization. Rare words are more important for characterizing differences across documents than common words. To take that into account, I weight each word using term-frequency-inverse-document-frequency (tf-idf). Each term m of the dataset is weighted by:

$$\omega_m = \log\left(\frac{K+1}{d_m+1}\right) + 1 \text{ where } d_m = \sum_k \mathbb{I}\left\{c_{km} > 0\right\}.$$

After weighting, each document is represented by word frequency vector f_k with entries:

$$f_{km} = \frac{\omega_m c_{km}}{\sqrt{\sum_{m'}(\omega_m c_{km})^2}}.$$

Singular Value Decomposition. Let F represent the entire corpus. Latent semantic analysis decomposes the matrix F as

$$F = U\Sigma V'$$

where U is the term-factor matrix, Σ the diagonal matrix of singular values, and V' the factor-document matrix.

Rank-k approximation. I set to zero all the singular values outside the top 300, following Schwarz (2019). Using the decomposition, I recover the matrix $\tilde{F}$ given by

$$\tilde{F} = U\tilde{\Sigma}V'$$

where $\tilde{\Sigma}$ has only the first 300 singular values different from zero.

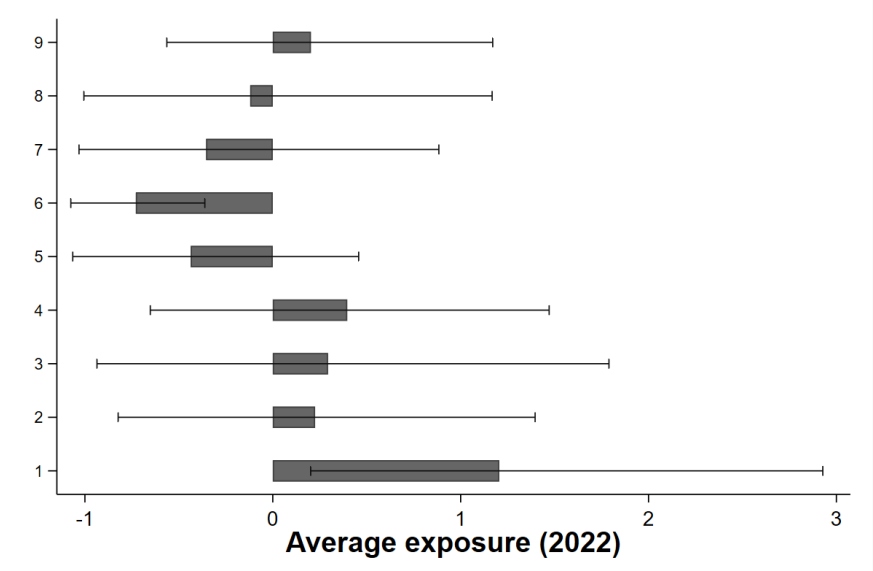
Similarity Scores. The similarity between software s and tasks of occupation o is given by:

$$s_{so} = \sum_{m \in \mathbb{M}} \tilde{f}_{sm} \times \tilde{f}_{om}. \tag{21}$$

C.2 Summary Statistics of AI Exposure

Figure A15 shows average AI exposure in 2022 across 1-digit occupation groups. Brackets indicate the range from the bottom to the top decile within each category. Exposure varies widely both across and within groups: managers, science professionals, and administrative workers are the most exposed, while agricultural workers are the least exposed.[34] Within-group variation is particularly notable: the most exposed manufacturing worker, for example, faces greater AI exposure than the average administrative worker. This wide distribution of AI exposure across occupations likely reflects its diverse range of applications, as discussed in Section 4.

Figure A15: **Large Variance within and Across Broad Occupations in AI Exposure**



Notes: This figure plots the average AI exposure in 2022 at a one-digit CBO occupational classification. The bars plot the average AI exposure and the lines the top and bottom decile.

Table A6, in the Appendix, shows the correlation between the 2003 characteristics of each occupation and their exposure to AI software in 2022. Occupations more exposed to AI

[34] Prytkova et al. (2024), Gmyrek et al. (2023), Felten et al. (2018), and Webb (2020) also find the same general patterns. In the Appendix Section C.3, I compare the AI exposure measure in Equation (5) to other measures developed in the literature.

tend to have higher hourly wages, greater education, and larger workforces. They are also more likely to involve tasks related to planning, management, and interpersonal activities, which is consistent with the finding that 40% of AI software targets administrative tasks.

Table A6: **Correlation Between AI Exposure and Characteristics of the Occupation**

	(1) $AI\ Exposure_{o,2022}$	(2) $AI\ Exposure_{o,2022}$	(3) $AI\ Exposure_{o,2022}$
$log(hour\ wage)_{o,2003}$	0.651^{***}	0.556^{***}	0.483^{***}
	(0.0555)	(0.0599)	(0.0627)
$log(yrs.\ educ.)_{o,2003}$	1.130^{***}	1.345^{***}	1.091^{***}
	(0.164)	(0.205)	(0.214)
$log(n.\ workers)_{o,2003}$	0.0531^{***}	0.0629^{***}	0.0735^{***}
	(0.0125)	(0.0130)	(0.0136)
$\mathbb{I}\{use\ computer\}_o$	0.162^{**}	0.241^{***}	0.222^{***}
	(0.0630)	(0.0650)	(0.0665)
$\mathbb{I}\{operate\ machine\}_o$	-0.0601	-0.0291	-0.0116
	(0.0626)	(0.0675)	(0.0676)
$\mathbb{I}\{Planning\ or\ Execution\}_o$	0.295^{***}	0.207^{***}	-0.00117
	(0.0757)	(0.0753)	(0.0762)
$\mathbb{I}\{Creation\ and\ Innovation\}_o$	-0.255^{**}	-0.249^{**}	-0.176^{**}
	(0.0837)	(0.0831)	(0.0823)
$\mathbb{I}\{Management\}_o$	0.333^{***}	0.282^{***}	0.217^{**}
	(0.0599)	(0.0600)	(0.0619)
$\mathbb{I}\{Interpersonal\}_o$	0.381^{***}	0.435^{***}	0.432^{***}
	(0.0635)	(0.0688)	(0.0726)
Controls		1-digit Occ. FE	2-digit Occ. FE
Observations	2,148	2,148	2,148
R^2	0.3809	0.4100	0.4731

Notes: This table reports regression results where the dependent variable is the stock cosine similarity in 2022. Each regressor is defined at the occupational level in 2003: *log(hour wage)* is the log of average hourly wage; *log(yrs. educ.)* is the log of average years of schooling; *log(n. workers)* is the log of the number of workers; $\mathbb{I}\{use\ computer\}$ is a dummy if among the tools used by the occupation is a computer; $\mathbb{I}\{operate\ machine\}$ is a dummy taking one if among the tasks performed by the occupation there is machine operation; $\mathbb{I}\{Planning\ or\ Execution\}$ is a dummy taking one if the occupation performs tasks related to planning or execution, $\mathbb{I}\{Creation\ and\ Innovation\}$ is a dummy taking one if the occupation performs tasks related to creation or innovation, $\mathbb{I}\{Management\}$ is a dummy if the occupation performs tasks related to management, and $\mathbb{I}\{Interpersonal\}$ is a dummy taking one if the occupation performs tasks related to interpersonal relationships. Standard errors in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

C.3 Comparison to existing exposure metrics

Table A7 shows that the exposure measure defined in Equation (5) is positively correlated ($\rho = 0.417$) with the GPT-4-based task replaceability scores in Gmyrek et al. (2023), but negatively correlated with both the feasibility scores of Frey and Osborne (2017) and the patent-similarity measure of Webb (2020). One possible explanation is that earlier metrics attempt to predict what AI could do—drawing on expert judgment, capability benchmarks, or patent abstracts—while the NIIP-based measure captures what AI is already doing in practice.

The low correlations across metrics highlight the lack of consensus in the literature on which occupations are most exposed to AI. The two most used metrics of AI exposure, those developed by Webb (2020) and Frey and Osborne (2017), are negatively correlated while the remaining pairwise correlation does not exceed 0.417.

Table A7: **Correlation Between Different AI Exposure Metrics**

Variables	(1)	(2)	(3)	(4)	(5)	(6)
(1) *AI Exposure*$_{o,2022}$	1.000					
(3) Felten et al. (2018)	0.198	1.000				
(4) Frey and Osborne (2017)	-0.338	-0.296	1.000			
(5) Gmyrek et al. (2023)	0.417	0.038	-0.188	1.000		
(6) Prytkova et al. (2024)	0.147	0.064	0.132	0.249	1.000	
(7) Webb (2020)	-0.178	0.189	0.167	-0.143	0.120	1.000

Notes: This table shows the correlation between the exposure measure in Equation (5), evaluated in 2022, the last year of the sample, and other exposure metrics developed in the literature. To calculate the correlation, I used crosswalks from their original occupational classifications to CBO 2002.

C.4 Software Development Time

To estimate the time required to develop software, I measure the interval between when a programmer is hired and when they are first credited as an author of a registered software. To construct this measure, I merge NIIP software registration records with RAIS data, linking each software entry to the firm that owns it and to the programmer who authored it. This produces a firm–worker panel from 2002 to 2016, restricted to software programmers. For each programmer–firm pair, I identify the initial hiring date, keeping only the first recorded employment spell to avoid ambiguity from repeated hires. The development time is then

calculated as the number of days between that hiring date and the publication of the worker's first registered software.

Figure A16: **Distribution of Time to Build Software in Years**

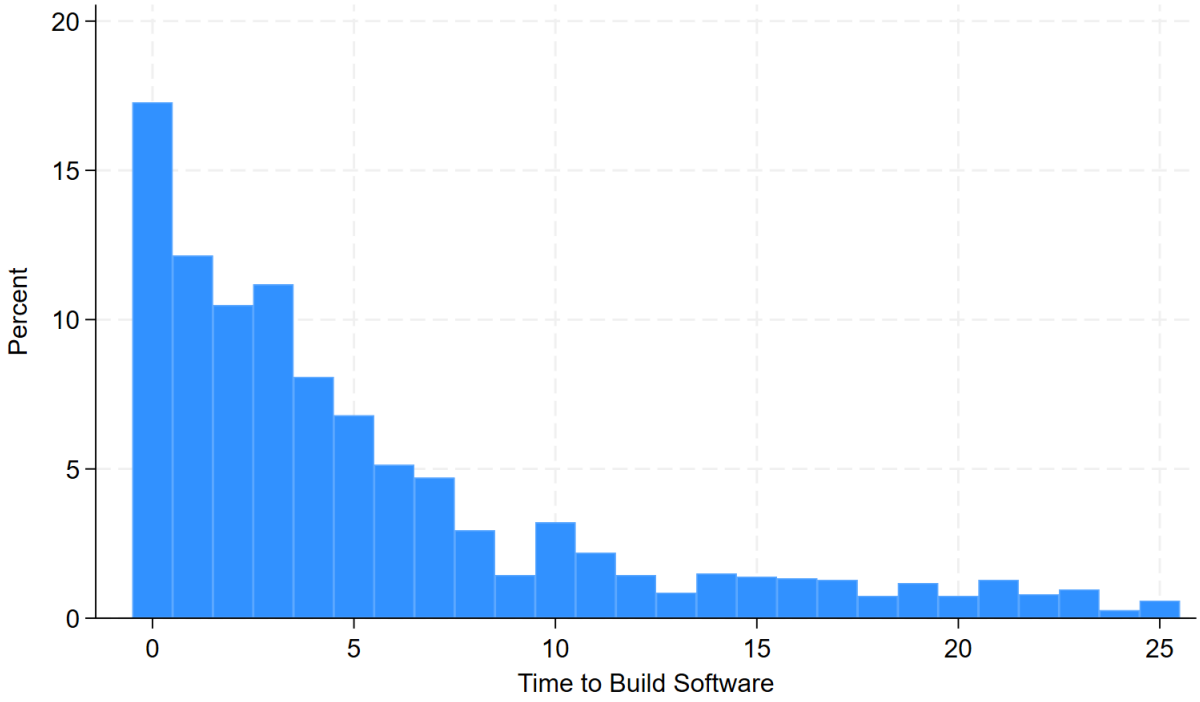


Table A8: **Time to Build Software**

	Years
Mean	5.4
Median	3
10th Percentile	0
25th Percentile	1
75th Percentile	7
90th Percentile	15

Notes: This table summarizes the distribution of time between hiring and the first software publication by a programmer. Time is measured in years.

C.5 Instrument Validation

Table A9: **Instrument and Exposure to Other Technologies**

	(1) *Non-AI Software*	(2) *Patents*	(3) *Industrial Design*
AI Ease-of-Dev IV	0.00815***	0.000276	0.00157
	(0.000754)	(0.000251)	(0.00102)
N	43,724	38,580	38,580

Notes: This table reports estimates from regressions of exposure to other technologies on the AI ease-of-development instrument. The dependent variables are exposure to non-AI software, patents, and industrial design, constructed using the same text-similarity approach as the AI exposure measure. Variables are normalized to have standard deviation one. Standard errors appear in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A10: **Instrument and Import Prices of Occupation-Related Equipment**

	(1) *All Products*	(2) *Capital Goods*	(3) *Machines*	(4) *High-Tech Equipment*
AI Ease-of-Dev IV	−0.0000740	0.000470	−0.000219	0.000359*
	(0.000121)	(0.000465)	(0.000140)	(0.000202)
N	43,724	43,724	43,724	43,724

Notes: This table reports estimates from regressions of occupation-level import price measures on the AI ease-of-development instrument. To construct the dependent variables, I link Harmonized System products to occupations using the text similarity between product descriptions and occupational tasks, then average import prices using these similarities as weights. Columns 2–4 restrict the products to capital goods, machines, and high-tech equipment. Standard errors appear in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A11: **Instrument and Trade Shocks**

	(1) *Input Import Prices*	(2) *Produced Imports*	(3) *Produced Export Prices*
AI Ease-of-Dev IV	0.00940	0.0263	−0.0166
	(0.00744)	(0.0309)	(0.0364)
N	45,362	42,294	42,277

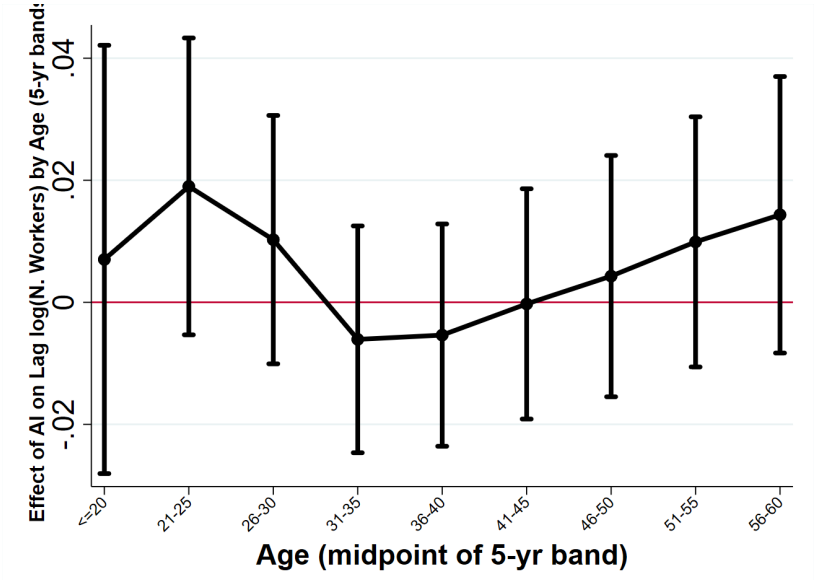
Notes: This table reports estimates from regressions of occupation-level trade exposure measures on the AI ease-of-development instrument. Column 1 uses import prices for input goods; column 2 uses imports of goods produced by sectors associated with the occupation; and column 3 uses export prices for produced goods. Occupation-level exposure to each trade measure is constructed using the sectoral distribution of employment within each occupation. Standard errors appear in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table A12: **Instrument and Lagged Labor Market Outcomes**

	(1) *log(Employment)*	(2) *log(Yrs. Educ.)*	(3) *log(HS Dropout)*	(4) *log(HS or More)*	(5) *log(Wage)*	(6) *log(Hourly Wage)*
AI Ease-of-Dev IV	−0.0118	−0.00204	−0.00488	−0.0106	−0.00121	−0.00244
	(0.00749)	(0.00140)	(0.0103)	(0.0117)	(0.00432)	(0.00569)
N	37,694	37,693	31,136	35,735	37,694	37,694

Notes: This table reports estimates from regressions of lagged labor market outcomes on the AI ease-of-development instrument. The dependent variables are measured two years before the current observation. Columns 1–6 report lagged log employment, average years of education, employment of high-school dropouts, employment of workers with high school or more, average wage, and hourly wage. Standard errors appear in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure A17: **Effect of AI on Lagged Employment of Different Age Groups**



Notes: This figure reports the effect of AI on employment by worker age group. Each point plots the estimated effect of AI on the log number of workers of a given age group. All specifications use the AI ease-of-development instrument defined in Equation (8). Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, exposure to U.S. AI patents, and the lagged outcome variable. Standard errors are clustered at the occupation level. Vertical bars report 95% confidence intervals.

C.6 Mincer Regression to Identify Abilities

To estimate individual ability, I run a Mincer regression using the version of RAIS with individual and firm identifiers. The dataset runs from 2003 to 2016, capturing only the initial years of the AI boom. The specification is given by

$$log(Wages_{i,t}) = \mu_i + X_{i,t} + \epsilon_{i,t}$$

where $Wages_{i,t}$ is monthly wages and μ_i is a worker fixed effect. $X_{i,t}$ is a set of controls including a 6-digit occupation fixed effect, an education-year fixed effect, an age fixed effect, a gender fixed effect, a 4-digit sector-year fixed effect, and a city-year fixed effect.

To calculate the average ability in occupation o in year t, used in Table 4, I average the fixed effect μ_i for workers in occupation o in year t. The wage residual in occupation o and year t is given by the average of $\epsilon_{i,t}$ for workers in that occupation in that year.

C.7 Task Group Construction

I construct task groups from the CBO 2002 task descriptions using the BERTopic pipeline (Grootendorst 2022). The starting point is the universe of 49,660 task descriptions. First, I represent each task as a sentence embedding using paraphrase-multilingual-MiniLM-L12-v2, a multilingual Sentence-BERT model (Reimers and Gurevych 2019). Sentence embeddings map each task into a vector space in which tasks with similar meanings are close to one another, even if they use different words. Second, I reduce these high-dimensional embeddings to five dimensions using UMAP, a nonlinear dimension-reduction method designed to preserve local neighborhoods in the data (McInnes et al. 2018). This step makes the clustering problem tractable while keeping nearby tasks close together. Third, I cluster the reduced embeddings using HDBSCAN, a density-based clustering algorithm (Campello et al. 2013, 2015). Unlike k-means, HDBSCAN does not require me to choose the number of clusters in advance and does not force every task into a group. Instead, it forms a group only when there is a sufficiently dense set of similar tasks; tasks in sparse regions are labeled as unassigned.

In the baseline classification, I require each HDBSCAN cluster to contain at least 30 tasks. This procedure yields 700 task groups and leaves 8,019 highly idiosyncratic tasks unassigned rather than forcing them into weakly related groups. I then name each task group using BERTopic's class-based TF-IDF procedure, or c-TF-IDF (Grootendorst 2022), which builds on the standard TF-IDF logic of giving high weight to words that are common in one document but rare in the broader corpus (Salton and Buckley 1988). In this setting, each task group is treated as a document. A word receives a high score if it appears frequently in that group but rarely in other groups. The resulting names are descriptive labels used to interpret the groups; they are not regression inputs. Finally, I map the task groups back to occupations and construct two occupation-level variables for each group: an indicator equal to one if the occupation contains at least one task in the group, and the share of the occupation's tasks that belong to the group.

C.8 Construction of the Information Use Measure

I construct an occupation-level measure of information use from the CBO task descriptions. The goal is to distinguish tasks in which workers use information to guide actions or decisions from tasks in which workers mainly process, record, classify, or organize information. The classification follows the same logic as the text-similarity approach in de Souza and Li (2023): I define reference texts for two conceptual categories, represent both tasks and reference texts in a common semantic space, and classify each task according to which reference category it is closer to.

First, I collect a set of benchmark activities describing tasks intensive in data use or intensive in data processing. The information-use benchmark uses the Wikipedia pages Instrumentação industrial, Controle estatístico de processos, Metrologia, Controle de qualidade, Telemetria, Manutenção preditiva, Agricultura de precisão, Monitoramento ambiental, Sensoriamento remoto, Sistema de posicionamento global, and Inteligência empresarial. The information-processing benchmark uses the Wikipedia pages Tratamento de dados, Processamento eletrônico de dados, Gerenciamento eletrônico de documentos, Banco de dados, Sistema de gerenciamento de banco de dados, Arquivística, Arquivo, Documento, Sistema de ficheiros, and Gestão de conteúdo empresarial.

I convert both the CBO task descriptions and the Wikipedia benchmark texts into sentence embeddings using the multilingual Sentence-BERT model(Reimers and Gurevych 2019, 2020). A sentence embedding maps each text into a numerical vector, with texts that have similar meanings located closer to one another. For each Wikipedia page, I split the page into text chunks, compute embeddings for the chunks, and average them to form one vector for the page. I then average the page vectors within each benchmark set to obtain two centroids: one for information use and one for information processing.

For each CBO task τ, I compute two cosine similarities: one between the task embedding and the information-use centroid, and one between the task embedding and the information-processing centroid. I define the task-level score as

$$Score_\tau = Sim(\tau, Use) - Sim(\tau, Processing).$$

A task is classified as information use if $Score_\tau > 0$, and as information processing otherwise.

I aggregate the task classifications to the occupation level. Let N_o be the number of

distinct tasks in occupation o. The information use intensity of occupation o is

$$\textit{Information Use Intensity}_o = \frac{1}{N_o} \sum_{\tau \in o} \mathbb{1}\{Score_\tau > 0\}.$$

C.9 Other Results

Table A13: **Stack Overflow Posts and AI Software Development**

	(1) *log(N. Software + 1)*	(2) *log(N. Software + 1)*	(3) *log(N. Software + 1)*	(4) *log(N. Software + 1)*
log(N. AI SO Posts)	0.205***	0.139***	0.194***	0.0565***
	(0.0138)	(0.0186)	(0.0136)	(0.0183)
log(N. Non-AI SO Posts)	0.168***	0.0296	0.186***	0.0852***
	(0.0101)	(0.0209)	(0.0101)	(0.0255)
N	2,503	2,446	2,503	2,446
R^2	0.331	0.795	0.357	0.822
F	617.6	60.37	663.7	23.70
Fixed Effects:				
Language	-	✓	-	✓
Year	-	-	✓	✓

Notes: This table reports estimates of the relationship between Stack Overflow posts and AI software development at the language-year level. The dependent variable is log number of AI software registrations plus one. The main regressors are the log number of AI and non-AI Stack Overflow posts by programming language and year. Standard errors appear in parentheses. $^{*}p < 0.10, ^{**}p < 0.05, ^{***}p < 0.01$.

Table A14: **Stack Overflow Posts and AI Software Development in Specialized Fields**

	(1) *log(N. Software + 1)*	(2) *log(N. Software + 1)*	(3) *log(N. Software + 1)*	(4) *log(N. Software + 1)*
log(N. AI SO Posts) × *AI Software Share*	0.877***	1.359***	0.942***	1.211***
	(0.00880)	(0.0186)	(0.0164)	(0.0237)
N	111,176	111,176	111,176	110,954
R^2	0.082	0.293	0.510	0.586
F	9,947.3	5,349.0	3,294.7	2,603.1
Fixed Effects:				
Language	-	✓	✓	-
Field	-	✓	✓	-
Year	-	-	✓	✓
Language-Field	-	-	-	✓

Notes: This table reports estimates of the relationship between Stack Overflow exposure and AI software development at the language-field-year level. The dependent variable is log number of AI software registrations plus one. The Stack Overflow exposure variable is the log number of AI Stack Overflow posts for a programming language interacted with the language's AI software share within an application field. Standard errors appear in parentheses. $^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

Table A15: **AI Increases Hiring and Firing**

	(1) *log(Hiring)*	(2) *log(Firing)*	(3) *Hiring Rate*	(4) *Firing Rate*
AI Exposure	0.0675***	0.0604***	0.0327*	0.0237
	(0.0148)	(0.0153)	(0.0184)	(0.0179)
N	49,961	49,560	50,441	50,441

Notes: This table reports 2SLS estimates of the effect of AI on hiring and firing, following the specification in Equation (6). Each column uses the AI ease-of-development instrument defined in Equation (8). Columns 1 and 2 use the log number of hires and firings; Columns 3 and 4 use hiring and firing rates. Controls include occupation fixed effects, year fixed effects, one-digit-occupation-by-year fixed effects, the lagged outcome, lagged non-AI software exposure, lagged exposure to U.S. AI patents, and the non-AI component of the ease-of-development instrument. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10$,** $p < 0.05$,*** $p < 0.01$.

Table A16: **Effect of AI by Gender, Age, Education, and Experience**

	(1)	(2)
Group	*log(Employment)*	*log(Wage)*
By age		
Young *(less than 35)*	0.0436***	0.00175
	(0.0103)	(0.00607)
Old *(more than 35)*	0.0175*	0.00347
	(0.0102)	(0.00537)
By education		
High School or more	0.0246**	0.00195
	(0.00959)	(0.00507)
High School incomplete	0.0288**	−0.00744
	(0.0142)	(0.00802)
By experience		
$>$ *one year*	0.0229**	0.00641
	(0.0104)	(0.00547)
$\leq$ *one year*	0.0740***	−0.0136*
	(0.0162)	(0.00820)
By gender		
Male	0.0306***	0.00691
	(0.00998)	(0.00543)
Female	0.0445***	0.00691
	(0.0126)	(0.00677)

Notes: This table reports estimates of the effect of AI on employment and wages across demographic groups, following the specification in Equation (6). All specifications use the AI ease-of-development instrument defined in Equation (8). For each specification, the left-hand side variable is calculated among the sample described in *Group* column. Column 1 reports the effect of AI on log employment; column 2 on the log of average wages. Controls include occupation fixed effects, one-digit-occupation-by-year fixed effects, exposure to non-AI software, and exposure to U.S. AI patents. Standard errors, clustered at the occupation level, appear in parentheses. $^{*}p < 0.10,^{**}p < 0.05,^{***}p < 0.01$.